Biomarkers for the response to immunotherapy in patients with non-small cell lung cancer

Mirte Muller

Biomarkers for the response to immunotherapy in patients with non-small cell lung cancer

Mirte Muller

Biomarkers for the response to immunotherapy in patients with non- small cell lung cancer

ISBN: 978-94-93330-86-3

The work described in this thesis was performed at the Netherlands Cancer-institute-Antoni van Leeuwenhoek, Amsterdam, the Netherlands

Author: Mirte Muller
Design and layout: Dagmar Versmoren, persoonlijkproefschrift.nl
Printed by Proefschriftspecialist | proefschriftspecialist.nl

Financial support for printing of this thesis was kindly provided by the Oncology Graduate School Amsterdam (OOA) and university of Leiden.

Biomarkers for the response to immunotherapy in patients with non-small cell lung cancer

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op donderdag 29 mei 2024

klokke 13:45 uur

door

Mirte Muller

geboren te Eindhoven

in 1988

Promotor:	Prof. Dr. P. Baas	
	Prof. Dr. M. M. Van den Heuvel	Radboud UMC
Copromotors	Dr. H.H. van Rossum	AvL-NKI
Leden promotiecommissie:	Prof. Dr. J.J.C. Neefjes	
	Prof. Dr. E.F. Smit	
	Prof. Dr. J.G. Borst	
	Prof. Dr. A.H. Maitland - van der Zee	AUMC
	Prof. Dr. Würdinger	AUMC

Table of content

	General introduction and outline of this thesis	8
Chapter 1	New targeted treatments for non-small cell lung cancer – role of nivolumab.	17
Chapter 2	Pembrolizumab for the treatment of non-small cell lung cancer.	37
Chapter 3	Blood platelet RNA profiles do not enable for nivolumab response prediction at baseline in patients with non-small-cell lung cancer.	55
Chapter 4	A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors.	69
Chapter 5	Prediction of Response to Anti-PD-1 Therapy in Patients with Non-Small Cell Lung Cancer by Electronic Nose Analysis of Exhaled Breath.	93
Chapter 6	eNose analysis for early immunotherapy response monitoring in non-small cell lung cancer.	115
Chapter 7	I love my dog, editorial.	139
Chapter 8	Validation of a clinical blood-based decision aid to guide immunotherapy treatment in patients with non-small cell lung cancer.	143
Chapter 9	Modelling Diagnostic Strategies to manage toxic adverse-events following cancer immunotherapy.	167
	General discussion and future perspectives.	188
Appendices	Nederlandse samenvatting	198
	Summary	201
	Dankwoord	204
	List of publications	206
	Curriculum Vitae	208
	References	209

Wat ik vroeger niet kon
Wat ik nu dus inmiddels wel kan

Introduction and thesis outline

Lung cancer is the leading cause of cancer death in the Netherlands, with 14.000 diagnoses and over 10.000 deaths per year in 2019 [11]. The major cause of lung cancer is smoking, which is responsible for 80% of cases in males and 50% of cases in females [12]. Non-small-cell lung cancer (NSCLC) accounts for 85–90% of all lung cancers. In the majority of cases, patients are diagnosed at an advanced, unresectable stage of disease [11]. For these patients, the treatment has a palliative intent, aiming to control symptoms and prolong survival.

In the last two decades, several novel therapeutic agents were developed, such as targeted treatment [13-16]. However, a targetable driver mutation is detectable only in 10%–20% of all NSCLCs in the Caucasian population [17]. For the others, chemotherapy was the only available option so far, with dismal results. In the last deaced, immune checkpoint inhibitors, such as nivolumab and pembrolizumab, have been shown to be highly active in different malignancies [18]. They bind with high affinity to the PD-1 receptor, an immune checkpoint, and promotes antitumor immunity and thus eliminate tumor cells [19, 20].

The first pilot study of nivolumab in heavily pretreated patients with non-small cell lung cancer (NSCLC) started in 2008 [7, 9]. This study, in which 76 patients with lung cancer were included, showed an overall response rate (ORR) of 19% (14 of 76 patients)[7]. Since then, immune checkpoint inhibitors have dramatically improved the treatment of advanced stage NSCLC.

The FDA and EMA approved nivolumab in 2015 for the treatment of pretreated advanced squamous cell carcinoma [21, 22]. However, due to its high costs per quality adjusted life year (QALY), which would be an estimated €130.000, nivolumab wasn't reimbursed as treatment for NSCLC in the Netherlands, despite the outstanding responses [23]. Therefore, while the negotiations with the Ministry of Health were ongoing, nivolumab was provided by Bristol-Myers Squibb (BMS) through a compassionate use program (Expanded Access Program, or EAP) from August 2015 for patients with advanced NSCLC. The results of this program from the patients who started their treatment in the Netherlands Cancer Institute are published by my colleague dr Schouten et al [24, 25].

Due to the high costs in combination with a relative low ORR, the urge to be able to identify responders or non-responders to immunotherapy was high. In other words: a cost-effective and accurate biomarker was needed. A biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biologic or pathogenic processes, or the response to a therapeutic intervention", as defined by Lee et al [26]. Criteria for a perfect biomarker are (1) understandable rationale, (2) Accurately predict or monitor a responder before treatment, (3) minimally invasive, (4) easy to collect and perform, (5) reproducible, robust and repeatable, (6) fast, and (7) cost-effective (table 1). For targeted treatment, their target (mostly a mutation), can be used as biomarker [15, 16]. For example, patients with an epidermal growth factor receptor (EGFR) mutation in

exon 19, show high sensitivity to erlotinib [16]. However, a biomarker for immunotherapy is far more challenging, since it involves the complexity of the immune system [27, 28].

Table 1 - The perfect biomarker

Requirement	Definition
1. Understandable rationale	A clear statement of reasons why a biomarker is specific (for immunotherapy). [29]
2. Accurately predict or monitor a responder before treatment.	The agreement between the best estimate of a quantity and its true value for the prediction of (non) response. [28, 30]
3. Minimally invasive	Not requiring a biopsy or other instrument in a part of a body (i.e. bronchoscopy or multiple blood samples). [29]
4. Easy to collect and perform	Not complicated to learn how to perform and collect a biomarker, with preferably as less steps as possible.
5. Reproducible, robust and repeatable	Measurements made under the same (repeatable) and different (reproducibility) conditions, including ambient temperature or storage condition of reagents [28, 30].
6. Fast	The time between sampling and result does not influence the time between (suspension of) diagnosis and start treatment.
7. Cost effective	Producing optimum results for the expenditure [29]

Biomarker development

Before a biomarker is used in clinical practice in actual patients, it needs to be validated (table 1). The important steps of the validation process are divided in three stages: (1) Analytical validation, (2) clinical validation and (3) clinical utility [26-28, 31]; preferably in a continuous process, which is also called the "fit-for-purpose" method [26]. Dr Lee et al [26], 2006, reported a conceptual strategy for the multiple steps for validation of a biomarker. They focus mainly on biomarkers for drugs, such as selection for treatment.

Analytical validation

Analytical validity is defined as the ability of a specific test to accurately and reliably measure the marker of interest in the clinical laboratory [28]. In short, it looks at the validation of marker itself (laboratory issues). When validated, a marker is performed in multiple certificated laboratories, which therefore needs parameters such as limit of detection, precision, and a reference range. The analytical validation can be divided into different steps, which are explained in short below [28].

Sample-related validation

This is the process before a biomarker is measured. This includes the use of a certain anticoagulants, the storage of a sample before use and the time between defrosting and processing. A standard operation protocol (SOP) is recommended and essential for the development of a biomarker. In here, background effect, such as hemolysis, should be evaluated [28] .

Assay-related validation

This includes the validation of things as precision, accuracy and robustness. The definition of precision is the agreement of repeatability (measurements made under the

same conditions) and reproducibility (measurements done under different conditions) [28, 30] and is a requirement for the implementation of all diagnostic tests [28]. For this, the use of negative and positive controls and a SOP are required. Definite quantitative assays make use of calibrators and a regression model to calculate absolute quantitative value for unknown samples. The reference standard should be well defined. Also good to remember: Precision can be validated; accuracy can only be estimated [28].

Data-related validation
This involves the interpretation of (continuous) assay results, in particularly in studies with 'big data'. In a case of RNA sequencing, this includes the algorithms for further data analysis [28] .

Clinical validation
Clinical validation shows the observation that a biomarker accurately predict the outcome of treatment, which is defined by the nature of the clinical question [27]. Internal validation would be the first step, with a training and validation set. A training set, a set with collected samples with a known outcome (retrospective), can be used to define, build, or calculate a test. The accuracy is presented in a receiver operating characteristics (ROC) curve. Here, the true positive rate (TPR, or sensitivity) and true negative rate (TNR, or 1-specificity) are plotted with a test with a continuous outcome. In a ROC curve the cut-off value can be determined for further investigation in the validation set. When determined, the test should be 'locked' and can not be changed. This predefined test is validated in an independent set of patients with a blinded but known outcome: the validation set. In a validation set the best test developed in the training set will be validated in a separate group of patients, not involved in the training group.

When these first results are promising, an external validation on an independed dataset should follow, preferably from another clinic. The number of patients of this independed validation cohort should be sufficient and calculated with a power calculation [27, 32].

Clinical utility
The clinical utility is a measure of whether clinical use of the test improves patient outcomes for a specific indication [27]. This should be validated in a prospective study, which show the benefit of the use of the marker. There are three well-known designs for the validation of a marker, namely the enrichment design, a stratified design and stratify design [27]. In the enrichment design, the biomarker is assessed before randomization. Only the patients with a positive biomarker are randomized between start (or stop) of treatment and a control group. For the stratified design, the biomarker is assessed before start treatment. Here, also patients with a negative biomarker are randomized between start (or stop) of treatment of the control group. In the strategy design patients randomize before the biomarker is assessed. If they are randomized in the "biomarker assessment group", they will be randomized in either start (or stop) treatment or a control group [27]. The chosen design depends on the related question.

Regularity requirements

There are two types of biomarker which can be approved by the FDA: companion diagnostics and complementary diagnostics. Companion diagnostics are " essential for the safe and effective use of a corresponding drug or biological product" [20, 33] and are designed for the use of a specific (group of) treatment and mostly validated in the same trial [34]. Approved drugs (or drug group) and their companion diagnostic refer to each other in their labels, as indicated in the FDA guidance [27]. Complementary diagnostics, where the biomarker is not a requirement for start (or stop) treatment, but provides useful information.

The FDA has established three different classes for the approval of medical devices, classified by risk. The definition of Class III by the FDA "Class III devices are those that support or sustain human life, are of substantial importance in preventing impairment of human health, or which present a potential, unreasonable risk of illness or injury" [35]. Both companion and complementary diagnostics are considered class 3. For class 3, besides the general and special controls, a premarket approval (PMA) is required [35]. A PMA is "a scientific, regulatory documentation to FDA to demonstrate the safety and effectiveness of the Class III device", as defined by the FDA, and consists of the technical, non-clinical laboratory study and a clinical investigation section [35].

PD-L1 as a biomarker

PD-L1 is a biomarker designed for the prediction on response to anti-PD-1 with expressing its target. High expression of PD-L1 in a biopsy of the tumor (primary or metastases), analyzed using immunohistochemistry (IHC), suggests a good response to immunotherapy [20]. Based on this hypothesis, biopsies were collected in the first pilot study of nivolumab, analyzed with PD-L1 IHC and compared with response [7, 9]. The first results were promising: PD-L1 positive showed an objective response rate (ORR) of 36% versus PD-L1 negative 0% [7]. However, in a bigger cohort of NSCLCs, no association was seen between PD-L1 status and either ORR or overall survival (OS) [9]. Across different trials, objective response (OR) and longer duration of response (DOR) have been registered both in PD-L1-positive and PD-L1-negative NSCLCs, even in numerically higher among positive tumors [2, 36] and no differences have been described for different levels of PD-L1 expression [2, 3, 5, 37]. As for PD-L1 and nivolumab, only a FDA approval for a complementary diagnostic is obtained[38].

The predictive role of PD-L1 expression is also investigated in trials evaluating pembrolizumab. In the KEYNOTE-001 study, PD-L1 expression was assessed in tumor samples using two assays.

Here, 23 out of 55 patients receiving 2 mg/kg Q3W pembrolizumab had PD-L1 expression in ≥50% of the tumor cells. Here, ORR was 30% (95% CI: 13–53%), which is relatively high.

In the Phase II/III study Keynote-010, PD-L1-positive (ie, PD-L1 expression ≥1% of tumor cells) NSCLC patients treated with pembrolizumab achieved a longer median OS compared to those receiving docetaxel (pembrolizumab 2 mg/kg: 10.4 months; pembrolizumab 10 mg/kg: 12.7 months; docetaxel 75 mg/m^2: 8.5 months). This survival advantage was higher for patients with ≥50% of PD-L1-positive tumor cells, despite the dose of pembrolizumab they received (HR 0.54 for pembrolizumab 2 mg/kg vs docetaxel,

95% CI 0.38–0.77, P=0.0002) [39]. In October 2015, the FDA approved pembrolizumab for the treatment of patients with metastatic NSCLC who are progressive on or after platinum-containing chemotherapy. This approval was based on the KEYNOTE-001 trial [40, 41]. In KEYNOTE-024, treatment naïve patients, with PD-L1 staining of at least 50%, the objective response rate to pembrolizumab was 44.8% [42]. This study provided the second FDA approval (in 2016) of PD-L1 for first line of treatment of NSCLC.

Despite these recommendations, it remains difficult to predict response using only PD-L1 expresssion. Other biomarkers are required for optimal selection. The research focusses to improve outcome of immunotheray besides PD-L1 as a biomarker. It has become clear that expression of other markers are tested like tumor mutational load and the tumor microenvirement [43]. Fumet et a. 2018 [44] showed that in patients with tumors with a high PD-L1 expression and a low count of cytotoxic T cell infiltrating lymfocytes (TILs), the OS was lower compared to those with a high count of TILs (3.7 months vs 8 months (p=0.02)). Other studies confirmed that a high TIL count in combination with a high PD-L1 expression resulted in a better prognosis [45-49].

Despite this research the single use of the PD-L1 marker is widely used, partly due to the FDA approval[50]. Having an FDA approval soon after the arrival of immunotherapy suggests that PD-L1 is an accurate and useful biomarker. It has good qualities, which meet some of the criteria for a perfect biomarker, as mentioned above. The rationale for PD-L1 is its clear: it directly shows the expression of the target of the immunotherapy. It is easy to perform, with some training of the pathologists. It is quite fast, a result can be expected within a week. The test is relatively cheap, more or less 100 euros[51]. However, although PD-L1 is able to select patients who would show more benefit from immunotherapy than others, the marker is not able to save treatment. A biopsy is mandatory for the evaluation of the PD-L1. Depending on the site of the tumor of the patient, this is at least a little invasive. Also, discordance between two biopsies from the same patient is a known phenomenon for the use of PD-L1 in clinical practice. Different studies show relatively high discordance scores for PD-L1 when using the three most used categories for PD-L1: tumor proportion score (TPS) <1%, 1-50% and >50%; with >50% being highly positive. For example, the discordance score between samples of the same biopsy lies between 23 and 59% [52-55]. In other words: in 23-59% of the two biopsies from the same side of tumor the PD-L1 category differed, which could lead to a different choice of treatment. The discordance between a primary and metastatic biopsy showed a score of 33-71.4% [52, 56, 57]. Another poor feature is that multiple assays may lead to discordant results while using the same biopsy, leading to another choice of treatment [20].

The main question is how many and which biomarkers are required to obtain a more reliable prediction of the success of immunotherapy. There are two hypotheses:

(1) It is possible to predict response with a surrogate biomarker.
(2) Are there more perfect biomarkers for the prediction of response to immunotherapy in patients with advanced NSCLC.

In the next chapters the search for any of more perfect biomarkers is described.

Chapter 1 and 2 are an introduction about immunotherapy. **Chapter 1** describes the standard of treatment in patients with NSCLC in 2015 and nivolumab as a new drug is introduced in terms of efficacy, safety and patients' quality of life. In **Chapter 2**, pembrolizumab is reviewed, where efficacy, safety, quality of life (QoL) and the role of PD-L1 testing are discussed. Both **Chapter 1 and 2** show the benefits of immunotherapy in terms of survival and toxicity. Nevertheless, results have shown that only a minority of patients experiences a relevant clinical benefit [58].

In 2015, the possibility to detect cancer with the use of platelets in one single drop of blood was investigated. Platelets contain fragments of RNA of the rest of the body, including tumor RNA, and thereby can support the tumor. For this reason they are called tumor educated platelets. Findings were published in Cancer Cell [59-61]. Our hypothesis was that these RNA fragments can predict whether someone will respond to immunotherapy. These findings are described in **Chapter 3.**

Another developed biomarker is the use of proteins in sera of patients treated with immunotherapy. An interesting theory, published by Blank et al [62], introduced the paradigm 'The Immunogram', in which proteins such as C-reactive protein (CRP) and lactate dehydrogenase (LDH) predict the response to immunotherapy [62]. Another publication showed that the ratio of CRP to albumin as a negative prognostic factor in both early and advanced NSCLC [63]. Also, complement activation in serum of patients with early-stage NSCLC was significantly associated with poor prognosis [64]. In **Chapter 4**, this knowledge is used for the development of a serum-based pretreatment protein test.

The biomarker that could be considered as least invasive is the one with the use of volatile organic compounds (VOCs) in exhaled breath. These VOCs originate from metabolic processes and thereby could tell something about the immunological responses. For our project an electronic nose, called the SpiroNose [65], was used. In **Chapter 5**, the accuracy of exhaled breath is researched in a group of patients with second line immunotherapy treatment at baseline. In **Chapter 6** this data is combined with follow up measurements of exhaled breath from 6 and 12 weeks after start of treatment.

Although the electronic nose shows promising results, as presented in chapter 5 and 6, the use of a dog outperforms (all) the developed electronic nose(s). **Chapter 7** is an intermezzo and this editorial outlines the pros and cons about the use of a dog.

Besides predicting, biomarkers can be used for monitoring of disease during treatment. For NSCLC follow-up several circulating tumor markers are available, however, most biomarkers in clinical practice have not been validated as monitoring tools. In **Chapter 8** the validation of a new clinical blood-based decision tool is discussed.

Whilst receiving treatment the aim is to gain a response on the one hand, while not getting a side effect. With immunotherapy there are less side effects compared to chemotherapy. However, when a side effect occurs, it is more likely to be severe. Therefore, patients are tested before every treatment cycle, to monitor the possibility of a side effect. In **Chapter 9**, these lab results, or markers, are used for the monitoring

of toxicity while receiving treatment. This modelling is one of the possibilities to combine different markers and optimize the use of them.

In the last chapter, the **general discussion and future perspectives**, the before mentioned chapters will be discussed in the light of being a biomarker now and in the future.

1 New targeted treatments for non-small cell lung cancer – role of nivolumab

G. Zago*, M. Muller*, M.M. Van den Heuvel, P. Baas.

Biologics 2016; 10: 103-117

*Contributed equally

Abstract

Non-small-cell lung cancer (NSCLC) is often diagnosed at an advanced stage of disease, where it is no longer amenable to curative treatment. During the last decades, the survival has only improved significantly for lung cancer patients who have tumors harboring a driver mutation. Therefore, there is a clear unmet need for effective therapies for patients with no mutation. Immunotherapy has emerged as an effective treatment for different cancer types. Nivolumab, a monoclonal inhibitory antibody against PD-1 receptor, can prolong survival of NSCLC patients, with a manageable toxicity profile. In two Phase III trials, nivolumab was compared to docetaxel in patients with, respectively, squamous (CheckMate 017) and non-squamous NSCLC (CheckMate 057). In both trials, nivolumab significantly reduced the risk of death compared to docetaxel (41% and 27% lower risk of death for squamous and non-squamous NSCLC, respectively). Therefore, nivolumab has been approved in the US and in Europe as second-line treatment for advanced NSCLC. Unfortunately, accurate predictive factors for patient selection are lacking, making it difficult to decide who will benefit and who will not. Currently, many ongoing trials evaluate the efficacy of nivolumab in different settings and in combination with other agents. This paper reviews the present literature about the role of nivolumab in the treatment of NSCLC. Particular attention has been given to efficacy studies, toxicity profile, and current and emerging predictive factors.

Keywords: advanced non-small-cell lung cancer; anti-PD-1; immunotherapy; nivolumab.

Introduction

Lung cancer is the leading cause of cancer death worldwide, with 1.825 million diagnoses and 1.59 million deaths in 2012 [66] and is the most commonly diagnosed malignancy in males. The major cause of lung cancer is smoking, which is responsible for 80% of cases in males and 50% of cases in females [12]. Non-small-cell lung cancer (NSCLC) accounts for ~85%–90% of all lung cancers. The two major subtypes are non-squamous cell (mainly adenocarcinoma) and squamous cell carcinomas [14, 67]. In the majority of cases, patients are diagnosed at an advanced, unresectable stage of disease. For these patients, the treatment has a palliative intent, aiming to control symptoms and prolong survival.

In the last decades, the discovery of genomic heterogeneity of NSCLCs has radically changed the diagnostic approach for these patients. With the advent of new techniques (integrating morphological analysis, immunohistochemistry, and molecular testing), different subclasses of NSCLCs have been defined (Figure 1) [68]. Targetable alterations are the key elements for personalized treatments and are nowadays part of the standard of care for NSCLC patients. However, a targetable driver mutation is detectable only in 10%–20% of all NSCLCs in the Caucasian population (Table 1) [17, 67, 68]. For the others, chemotherapy has been the only available option so far, with dismal results.

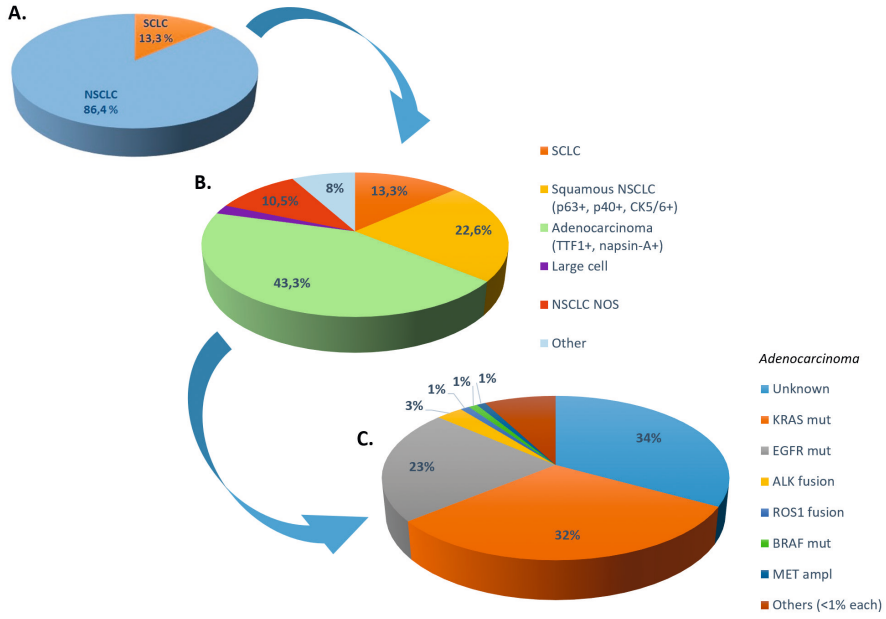


Figure 1 - Multistep process for the diagnosis and characterization of lung cancer.
Notes: *(A) The two main lung cancer subtypes, SCLC and NSCLC, can be discriminated by a morphological analysis. NSCLC accounts for ~85%–90% of all lung cancers. (B) Immunohistochemistry allows different NSCLC subtypes to be distinguished. (C) Molecular testing allows possible driver mutations in the tumor to be identified (EGFR and ALK). Analysis of ROS1, BRAF, and MET should be considered for selected patients. Data from National Cancer Institute [73] and Naidoo et al [5].*
Abbreviations: *SCLC, small-cell lung cancer; NSCLC, non-small-cell lung cancer; NOS, non-squamous.*

In the last years, new agents have been developed which enhance the host immune response against the tumor. Immune checkpoint inhibitors have been shown to be highly active in different malignancies [18]. The anti-PD-1 monoclonal antibody (mAb) nivolumab has recently been shown to induce a significant survival benefit in NSCLC patients, with either squamous or non-squamous histology, compared to standard second-line chemotherapy [3, 5] thus providing a new treatment option in this setting.

The aim of this paper is to present the standard of treatment and review the clinical data about the role of nivolumab in the treatment of NSCLC, in terms of efficacy, safety, and patients' quality of life (QoL) [69-71].

Current and emerging treatment options for NSCLC

Until recently, chemotherapy has been the only available option for patients diagnosed with NSCLC not amenable to radical-intent local treatment. First-line chemotherapy doublet regimens, based on platinum compounds (cisplatin or carboplatin) combined with a third-generation agent (vinorelbine, gemcitabine, paclitaxel, docetaxel, pemetrexed), prolong survival and improve QoL. Mono-chemotherapy is considered the treatment of choice both as second line and for unfit or elderly patients [14]. In the last decade, with the discovery of driver mutations in a variable percentage of NSCLCs (mainly in never smokers or light former smokers, with a non-squamous histology), targeted therapies have emerged as new standard of care in this setting (Table 1). For tumors with an activating mutation, in the HER domain, EGFR-tyrosine kinase inhibitors, gefitinib, erlotinib, and afatinib, have shown higher efficacy, in terms of response rate (RR) and progression-free survival (PFS), compared to chemotherapy [72-75]. Similar results have been achieved by the ALK-tyrosine kinase inhibitors crizotinib and ceritinib, in tumors with ALK rearrangement [76-78]. More recently, new primary or acquired targetable molecular alterations have been identified, such as ROS1 rearrangement, MET amplification, and HER2 mutation. A number of Phase I and II trials have shown encouraging results (Table 1) [78-87], so molecular testing is now recommended for these genes in selected patients.

Recently, new immune-modulating drugs have been developed which target different immune checkpoints, with the aim of enhancing the host immune response against tumor cells. For patients with NSCLC, the best results have been shown by PD-1/PD-L1 immune checkpoint inhibitors, which might have a higher activity in high-mutational load tumors [88]. Nivolumab (Opdivo®; Bristol-Myers Squibb, New York, NY, USA; other names: BMS-936558, MDX-1106, ONO-4538), a PD-1-blocking antibody, has been approved by the US Food and Drug Administration (FDA) as second-line treatment for squamous NSCLC in March 2015. In October 2015, the FDA expanded its approved use to all NSCLCs (both squamous and non-squamous histology) that have progressed after a first-line platinum-based chemotherapy [3, 5].

Table 1 - Driver mutations, and current and emerging targeted treatments in NSCLC.

Molecular alteration	Frequency	Targeted agent	Studies and findings
EGFR mut (exon 19-21)	*Caucasian pts*: 10-15% NSCLCs *Asian pts*: 50% NSCLCs (Common mut: *Ex 19 Del*: 45%; *Ex 21 L858R*: 40%)	***Gefitinib*** ***Erlotinib*** ***Afatinib***	IPASS [72]: 1st line Gefitinib is superior to Carbo-Paclitaxel in terms of RR (71% *vs* 47%) and PFS (9.5m *vs* 6.3m) in Asian, non-smoker pts, with ADC NEJSG002 [89]: 1st line Gefitinib is superior to Carbo-Paclitaxel in terms of RR (73% *vs* 30%) and PFS (10.8m *vs* 5.4m) in EGFR+ ADC EURTAC [73]: 1st line Erlotinib is superior to platinum-based ChT for RR (55% *vs* 11%) and PFS (9.7m *vs* 5.2m) in Caucasian pts with EGFR+ ADC LUX-Lung 3[80]: 1st line Afatinib is superior to Cis-Pem for RR (56% *vs* 23%) and PFS (11m *vs* 6.9m) in EGFR+ ADC LUX-Lung 6 [75]: 1st line Afatinib is superior to Cis-Gem for RR (67% *vs* 23%) and PFS (11m *vs* 5.6m) in EGFR+ ADC Pooled analysis LUX-Lung 3/6 [90]: 1st line Afatinib improves OS (31.7m *vs* 20.7m) for EGFR ex19del ADC compared to platinum-based ChT
EGFR mut (ex 20 T790M)	*Acquired resistance to 1st line EGFR TKi*: 50% pts	***AZD9291*** ***CO-1686*** ***(Rociletinib)***	AZD9291 phase I[79] : AZD9291 is effective in pts with T790M+ and T790M neg ADC, after a previous TKi treatment, with DCR=84% (*T790M+*: 95%; *T790M neg*: 61%) and median PFS=8.2m (*T790M+*: 9.6m; *T790M neg*: 2.8m) CO-1686 phase I/II [80]: Rociletinib achieves high DCR in pretreated, caucasian, T790M+ (93%) and T790M neg (59%) ADC pts, with median PFS 13.1m for T790M+ pts (80% pts censored)
ALK translocation	2-7% NSCLCs	***Crizotinib*** ***Ceritinib*** ***Alectinib***	PROFILE 1007 [76]: Crizotinib is superior to Pem *or* Docetaxel as 2nd line therapy in ALK+ NSCLC, in terms of RR (65% *vs* 20%) and PFS (7.7m *vs* 3m) PROFILE 1014 [77]: Crizotinib is superior to platinum-Pem as 1st line therapy in ALK+ NSCLC, in terms of RR (74% *vs* 45%) and PFS (10.9m *vs* 7m) ASCEND-1 phase I[78]: Ceritinib is effective for ALK+ NSCLCs, both pre-treated with or naïve to Crizotinib (RR=58%; PFS=7m) AF-001JP phase I/II [81]: Alectinib is effective for the treatment of ALK+ Crizotinib naïve pts (RR=93.5%; PFS=27.7m) AF-002JG phase I/II [82]: Alectinib is effective for the treatment of ALK+ Crizotinib-resistant pts (RR=55%; PFS N/R)
ROS1 rearrangement	1-2% NSCLCs	***Crizotinib***	Phase I [83]: Crizotinib achieves 72% ORR in pts with ROS1 rearranged NSCLC; estimated median duration of response: 17.6 m; median PFS: 19.2 m
MET amplification	<1% ADC	***Crizotinib***	Ongoing phase I trial NCT00585195 (PROFILE 1001)[84]: 13 pts (low/intermediate/high level of amplification) 4 PR (mainly highly amplified NSCLCs); 6 ongoing at data-cutoff Diarrhea (50%), nausea and vomiting (31%), peripheral edema (25%)

Table 1 - Driver mutations, and current and emerging targeted treatments in NSCLC. *(Continued)*

Molecular alteration	Frequency	Targeted agent	Studies and findings
BRAF V600E mutation	<2% NSCLCs	***Dabrafenib*** ***Dabrafenib + Trametinib*** *(MEK-inhib.)*	Ongoing phase II trial NCT01336634[85]: 84 pts in total (78 pretreated; 6 treatment naïve) Pretreated pts (n=78): 25 PR (ORR: 32%); DCR>12weeks: 56%; median duration of response: 11.8 months (95% CI, 5.4-N/R) Naïve pts (n=6): 3 PR (4 pts evaluable for response) Ongoing phase II trial NCT01336634[86]: 33 pts (24 pts evaluable for response) Response (n=24): 15 PR (ORR 63%); DCR>12weeks: 88% Common AEs (>20%): pyrexia, diarrhea, nausea, vomiting, peripheral edema, rash. Grade 3-4 AEs: 39% pts
HER2 mutation	3% ADC (2% NSCLCs)	***Trastuzumab*** *or* ***Afatinib***	Retrospective [87]: 65 pts HER2 ex 20 mut; 33 pts received HER2-targeted therapy ORR=56%; overall DCR: 82% (*Trastuzumab*: 96%; *Afatinib*: 100%) PFS for HER2-targeted therapies: 5.1 months
RET rearrangement	1-2% ADC	***Cabozantinib***	Ongoing phase II trial NCT01639508 [91] *:* 5 pts: 3 evaluable for response to treatment with Cabozantinib Early and durable response

Note: *EGFR+ represents EGFR mutation positive (ie, activating mutations)*
Abbreviations: NSCLC, non-small-cell lung cancer; pts, patients; RR, response rate; PFS, progression-free survival; ADC, adenocarcinoma; ChT, chemotherapy; Cis, cisplatin; Pem, pemetrexed; OS, overall survival; TKI, tyrosine kinase inhibitor; DCR, disease control rate; N/R, not reported; ORR, overall response rate; PR, partial response; CI, confidence interval; AEs, adverse events.

Pharmacology of nivolumab

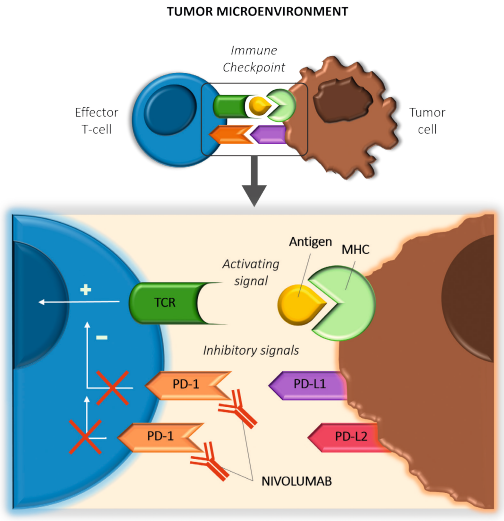


Figure 2 - Immuno-modulatory role of PD-1 receptor and mechanism of action of nivolumab.
Abbreviations: MHC, major histocompatibility complex; TCR, T-Cell Receptor; PD-1, programmed death 1; PD-L1, programmed death 1 – ligand 1; PD-L2, programmed death 1 – ligand 2.

Mechanism of action and pharmacodynamics

Nivolumab is a fully human IgG4 immune checkpoint inhibitory antibody, which binds to PD-1, preventing its interaction with its ligands PD-L1 (also called B7-H1 or CD274) and PD-L2 (also called B7-DC or CD273) (**Figure 2**) [92, 93]. PD-1 is an immune-regulatory receptor expressed by activated T-cells, and it is induced during any inflammatory reaction. PD-1 is also highly expressed by $CD4^+$ regulatory T (Treg) cells, and its main role is to limit immune response and maintain immune tolerance within peripheral tissues [94, 95], by both limiting the activity of effector T-cells and enhancing activity of inhibitory Treg cells [92]. Therefore, the interaction of nivolumab with PD-1 attenuates the negative signals of PD-1/PD-L1, thus enhancing the host antitumor immune response. Tumor infiltrating lymphocytes have been found to express PD-1 in different cancer types, while the upregulation of PD-L1 by tumor cells has been interpreted as a possible mechanism of resistance of the tumor to the host immune response [92, 96]. Lastly, PD-1-deficient mice showed a mild, late-onset immunological phenotype compared to the CTLA4-deficient ones, suggesting a more tolerable toxicity profile. These observations, taken all together, provided the basis for starting the development of new PD-1-targeted immunomodulatory compounds [18]. Preclinical studies have demonstrated that antibody-mediated PD-1/PD-L1 pathway blockade leads to an increase in T-cell count (both effector and antigen-specific) and modulates cytokines secretion in vitro and in murine models [97, 98]. The inhibition of PD-1 interaction with its ligand PD-L1, by specific mAbs, was able to rescue cytolytic immune antitumor activity, leading to tumor regression in mice [96].

The interaction of nivolumab with PD-1 receptor was evaluated using purified human T-cells from peripheral blood [10]. Nivolumab binds with high affinity to PD-1 on effector and memory T-cells and on Treg cells [10], thus preventing its interaction with PD-L1 and PD-L2. Neither $CD4^{+-}$ nor $CD8^{+-}$naïve T-cells are bound by nivolumab [10], reflecting the pattern of expression of PD-1, which is upregulated in activated T-cells in peripheral tissues [92]. Preclinical data showed that nivolumab binds to PD-1 on activated human $CD4^+$ T-cells with a half-maximal effective concentration (EC_{50}) of 0.64 nmol/L and inhibits PD-1 interaction with its ligands (PD-L1 and PD-L2) with a half-maximal inhibitory concentration (IC_{50}) of 2.52 and 2.59 nmol/L, respectively[10]. The high affinity of nivolumab for PD-1 has been confirmed in the first-in-human Phase I study by the analysis of PD-1 occupancy on circulating T-cells, which was demonstrated to be dose independent. Mean peak occupancy was 85% (range: 70%–97%), and the mean plateau occupancy was 72% (range: 59%–81%), detected after 4 hours to >57 days after the first infusion [1]. Plateau occupancy was maintained even when serum levels were undetectable [1]. Repeated infusions of nivolumab 10 mg/kg led to troughs and peaks of PD-1 occupancy around each dose, but 100% occupancy was not achieved [1].

Pharmacokinetics

Nivolumab is administered by intravenous infusion within 30–60 minutes. Pharmacokinetic data from the first-in-human Phase I trial demonstrated a half-life ranging between 12 (for subjects receiving the lowest doses: 0.3, 1, or 3 mg/kg) and 20 days (for subjects receiving the highest dose of 10 mg/kg), with a maximum serum

concentration directly related to administered dose within 4 hours after administration [1, 7]. No maximum dose has been defined [1, 7].

Nivolumab and NSCLC

Efficacy studies

Preclinical studies showed that nivolumab/PD-1 interaction leads to enhanced T-cell reactivity in vitro, in the presence of an antigen or another T-cell receptor stimulus [10]. Moreover, mAb inhibition of PD-1/PD-L1 interaction was able to rescue a cytolytic immune antitumor activity and lead to tumor regression in mice [96].

A Phase I trial with expansion cohorts was conducted between 2008 and 2012 (Figure 3 and Table 2), aiming to assess activity and safety of biweekly nivolumab at a dose of 1–10 mg/kg. A total of 296 heavily pretreated patients with advanced tumor were enrolled, including melanoma, NSCLC, renal cell carcinoma, and prostate and colorectal cancer. About one in four to one in five patients experienced durable objective response (OR), in particular those with melanoma, NSCLC, and renal cell cancer [7]. Given the encouraging results observed for NSCLCs, both in terms of RR and duration of response (DOR), an efficacy analysis based on data from a prolonged follow-up was carried out for this group of patients. In total, 129 patients with NSCLC underwent a median follow-up of 39 months (range: 32–66 months) and were evaluated for overall survival (OS), RR, and DOR [9]. Overall, median OS was 9.9 months (95% confidence interval [CI] 7.8–12.4), with a 1-, 2-, and 3-year survival rate of 42%, 24%, and 18%, respectively, without significant differences between squamous and non-squamous histologies. OS was longer for patients receiving nivolumab at the dose of 3 mg/kg (median OS =14.9 months, 95% CI 7.3–30.3), which was the dose of choice for the subsequent Phase III trials. Across all doses, overall response rate (ORR) was 17%, for both histological subtypes, but was higher for patients receiving 3 or 10 mg/kg doses compared to 1 mg/kg. Median estimated DOR for responders was 17 months, and an additional 10% of patientsexperienced long-lasting stability of disease (ie, stable disease [SD] ≥24 weeks) [9].

Given these encouraging results, a Phase II trial (CheckMate 063) was conducted between 2012 and 2013 (Figure 3 and Table 2) to investigate the efficacy of biweekly nivolumab 3 mg/kg, in patients with advanced squamous NSCLCs that had progressed after at least one platinum-based chemotherapy regimen and one more subsequent line of treatment. The efficacy of nivolumab was confirmed in this highly refractory group of patients, with 14.5% subjects achieving a partial response (PR) and 26% SD, with a high proportion of patients experiencing a durable response, in both groups [2]. Updated survival data were presented during the 2015 World Conference on Lung Cancer (WCLC) [37]. Out of 17 patients who achieved a PR, 13 (76%) had ongoing responses; thus, median DOR was not reached. Overall, median OS was 8.2 months (95% CI 6.1–10.9), and 1-year OS rate was 41%. Two Phase II Japanese trials achieved similar results in terms of ORR. Nivolumab-treated squamous NSCLC (n=35) showed an ORR of 25.7% (95% CI 14.2–42.1), while the ORR of non-squamous NSCLC (n=76) was 19.7% (95% CI 12.3–30.0) [99].

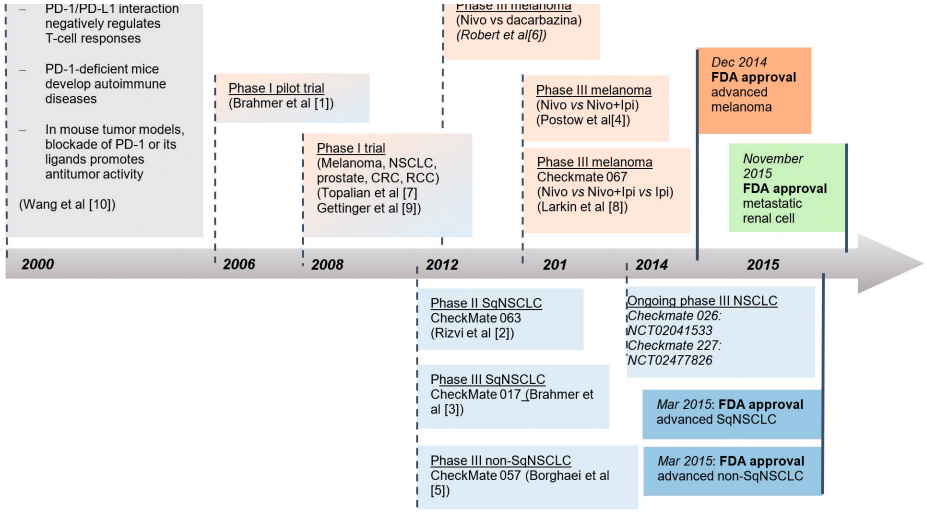


Figure 3 - Nivolumab development, from preclinical experience to clinical approval: focus on NSCLC.

Notes: *Timeline of nivolumab development from the preclinical studies to US FDA approval (dotted lines represent the starting date of the related trial).*

Abbreviations: PD-1, programmed death 1; PD-L1, programmed death 1 – ligand 1; *NSCLC, non-small-cell lung cancer; CRC, colorectal cancer; RCC, renal cell carcinoma; Nivo, nivolumab; Ipi, ipilimumab; FDA, US Food and Drug Administration; SqNSCLC, squamous NSCLC.*

The results achieved in the Phase II trial were confirmed by the two subsequent Phase III trials of second-line treatment nivolumab (**Figure 3**). CheckMate 017 and CheckMate 057 aimed to compare second-line treatment nivolumab (given biweekly at the dose of 3 mg/kg) to standard-of-care docetaxel monotherapy (75 mg/m^2, every 3 weeks), in advanced squamous (272 patients) and non-squamous (582 patients) NSCLC patients, respectively[8, 9]. In both trials, nivolumab significantly reduced the risk of death compared to docetaxel (41% and 27% lower risk of death for squamous and non-squamous histologies, respectively). For patients with squamous NSCLC, median OS was 9.2 months (95% CI 7.3–13.3) in the nivolumab group and 6 months (95% CI 5.1–7.3) in the control group, while for non-squamous NSCLC, OS was 12.2 months (95% CI 9.7–15.0) with nivolumab and 9.4 months (95% CI 8.1–10.7) with docetaxel. At 18 months of analysis, the OS rate was 28% vs 13% for squamous and 39% vs 23% for non-squamous carcinoma. Moreover, subgroup analysis from CheckMate 057 identified a lack of survival benefit with nivolumab in never smokers and EGFR-mutated tumors, albeit a small patient population [5]. This is in line with the Phase I trial, where nivolumab achieved a higher RR among current and former smokers compared to never smokers [9]. Both observations may find an explanation in the higher mutational load of smoking-induced tumors, which can lead to the production of a higher number of tumor-associated neo-antigens [100].

The results of previous trials have been recently confirmed in a large (n=824) ongoing study (CheckMate 153) conducted in community-based oncology centers. Among patients with advanced pretreated NSCLC, until now, no differences have been reported in terms of safety [101]. Among the 395 patients evaluable for tumor response, 55 (14%)

experienced a PR, and 194 (49%), a SD. No differences have been observed according to PD-L1 status or baseline performance status[101].

A number of trials are currently ongoing (**Table 2** and **Figure 3**) evaluating the role of nivolumab (alone or in combination) as first-line treatment for advanced NSCLC. Preliminary results from the Phase I trial CheckMate 012 (ClinicalTrials.gov: NCT01454102) have been presented during the 2015 American Society of Clinical Oncology (ASCO) Annual Meeting. In the cohort of the 52 chemotherapy-naïve patients who received nivolumab monotherapy, an ORR of 21%, with long DOR (median DOR not reached; range: 7.6+, 85.6+ weeks), was reported [36]. A Phase III trial (CheckMate 026, ClinicalTrials.gov: NCT02041533), comparing the first-line nivolumab to investigator's choice chemotherapy, in PD-L1-positive NSCLC, is currently ongoing [102].

During the 2015 WCLC, a manageable toxicity profile of nivolumab in combination with the CTLA4 immune checkpoint inhibitor ipilimumab (Yervoy™; Bristol-Myers Squibb; other names: MDX-010, MDX-101), in patients with NSCLC was reported. The ORR was 13%–39% across the four cohorts treated with different nivolumab and ipilimumab doses, but higher partial RRs were seen among patients who received nivolumab 3 mg/kg, and the median PFS was 4.9–10.6 months [103]. Thus, given also the positive results achieved by the same combination in advanced melanoma patients [4, 8], a Phase III trial is currently ongoing (CheckMate 227, ClinicalTrials.gov: NCT02477826) aiming to evaluate the OS of NSCLC patients receiving first-line nivolumab monotherapy, or nivolumab in combination with ipilimumab or chemotherapy, versus chemotherapy alone.

Other anti-PD1 and anti-PD-L1 compounds are currently under investigation as single agents or in combination for the treatment of NSCLC. Among the anti-PD-1 compounds, pembrolizumab (Keytruda®; Merck Sharp & Dohme Corp., Kenilworth, NJ, USA) was demonstrated to prolong OS compared to docetaxel, either at the dose of 2 (HR 0.71, 95% CI 0.58–0.88, P=0.0008) or 10 mg/kg (HR 0.61, 95% CI 0.49–0.75, P<0.0001) [51]. It has been approved by the FDA for the treatment of advanced NSCLC in October 2015. Results from Phase I and II trials are currently available about the role of the anti-PD-L1 compounds atezolizumab (or MPDL3280A) and durvalumab (or MEDI4736) in the treatment of NSCLC. In particular, results from the randomized Phase II trial POPLAR showed longer OS for atezolizumab compared to docetaxel (HR 0.73, 95% CI 0.53–0.99, P=0.04), in NSCLC patients after failure of a first-line platinum-based therapy. Moreover, the OS improvement correlated with PD-L1 expression[104]. No trials are currently available comparing efficacy and safety of these compounds.

Table 2 - Main trials evaluating nivolumab in NSCLC patients, from Phase I to Phase III trials, and preliminary results of ongoing trials.

Trial	Phase and line	Treatment	Patients	Findings
Gettinger et al [9]	Phase I Pretreated NSCLC	Nivolumab, q2w: - 1 mg/kg - 3 mg/k - 10 mg/kg	129 pts	ORR: 17%. ORR by dose: 3% (Nivo 1mg/kg), 24% (Nivo 3mg/kg) and 20% (Nivo 10gm/kg) Higher ORR in heavy smokers (>5 pack/years) Estimated duration of response: 17 m. Long lasting SD (≥24weeks): 10% Median OS: 9.9 months (95% CI, 7.8 to 12.4); Nivo 3 mg/Kg OS: 14.9 m; Nivo 1- or 10-mg/Kg OS: 9.2 m
Rizvi et al [2] CheckMate 063	Phase II 3rd line SqNSCLC	Nivolumab 3 mg/kg, q2w	117 pts	Response: PR=14.5% (95% CI 8.7-22.2); SD=26% (95% CI, 18-35) Median duration of response: not reached (95% CI, 8.3 m – N/R); median duration of SD: 6 m (4.7-10.6) Median OS: 8.2 m (95% CI, 6.1-10.9)
Brahmer et al[3] CheckMate 017	Phase III 2nd line SqNSCLC	Nivolumab *vs* Docetaxel	135 pts *vs* 137 pts	ORR: Nivolumab 20% (95% CI, 14-28) *vs* Docetaxel 9% (95% CI, 5-15) (p=0.008) Median OS: Nivolumab 9.2 m (95% CI, 7.3-13.3) *vs* Docetaxel 6 m (95% CI, 5.1-7.3) Risk of death 41% lower with Nivolumab (HR=0.59; 95% CI, 0.44-0.79; p<0.001)
Borghaei et al [5] CheckMate 057	Phase III 2nd line Non-SqNSCLC	Nivolumab *vs* Docetaxel	287 pts *vs* 268 pts	ORR: Nivolumab 19% (95% CI, 15-24) *vs* Docetaxel 12% (95% CI, 9-17) (p=0.02) Median OS: Nivo 12.9 m (95% CI, 9.7-15) *vs* Docetaxel 9.4 m (95% CI, 8.1-10.7) Risk of death 27% lower with Nivolumab (HR=0.73; 95% CI, 0.59-0.89; p=0.002)
Gettinger et al [36] CheckMate 012	Phase I 1st line NSCLC	Nivolumab 3 mg/kg, q2w	52 pts	*Ongoing*. Clin Trial Gov: NCT01454102 (*CheckMate 012*) *[Safety study of Nivolumab in combination with Cis/Gem, Cis/Pem, Carbo/Paclitaxel, Bevacizumab maintenance, Erlotinib, Ipilimumab or as monotherapy in pts with stage IIIB/IV NSCLC]* Nivolumab cohort: ORR=21%. Median duration of response: not reached (NR; range, 7.6+, 85.6+ weeks) Median OS: 98.3 weeks (range, 1.0-104.4+)

Table 2 - Main trials evaluating nivolumab in NSCLC patients, from Phase I to Phase III trials, and preliminary results of ongoing trials. *(Continued)*

Trial	Phase and line	Treatment	Patients	Findings
Rizvi et al [103] Checkmate 012	Phase I 1st line NSCLC	Nivolumab + Ipilimumab *(multiple doses)*	148 pts *(4 cohorts)*	*Ongoing*. Clin Trial Gov: NCT01454102 (*CheckMate 012*) *[Safety study of Nivolumab in combination with Cis/Gem, Cis/ Pem, Carbo/Paclitaxel, Bevacizumab maintenance, Erlotinib, Ipilimumab or as monotherapy in pts with stage IIIB/ IV NSCLC]* Nivo+Ipi cohort: ORR: 13–39%; median PFS: 4.9-10.6 m
Carbone [102]	Phase III 1st line PDL1+ NSCLC	Nivolumab *vs* ICC	Est. tot: 535 pts	*Ongoing*. Clin Trial Gov: NCT02041533 (*CheckMate 026*) Primary objective: PFS with Nivolumab *vs* ICC in pts with strong PD-L1 expression
ClinTrial.gov NCT02477826 CheckMate 227	Phase III 1st line NSCLC	Nivolumab *vs* Nivo+Ipi *vs* Nivo+ChT *vs* ChT	Est. tot: 1980 pts	*Ongoing*. Clin Trial Gov: NCT02477826 *(CheckMate 227)* An open-label, trial of Nivolumab, or Nivolumab plus Ipilimumab, or Nivolumab plus platinum-doublet Chemotherapy *vs* platinum doublet chemotherapy in subjects with stage IV NSCLC

Notes: Docetaxel dose in Phase III trials was 75 mg/m^2*; if not otherwise specified, nivolumab dose should be intended as 3 mg/kg, q2w.*

Abbreviations: NSCLC, non-small-cell lung cancer; CRC, colorectal cancer; RCC, renal cell carcinoma; Ipi, ipilimumab; SqNSCLC, squamous NSCLC; ORR, overall response rate; SD, stable disease; CI, confidence interval; OS, overall survival; PR, partial response; N/R, not reported; HR, hazard ratio; PFS, progression-free survival; ICC, investigator's choice chemotherapy; ChT, chemotherapy; q2w, biweekly; Cis, cisplatin; Gem, gemcitabine; Pem, pemetrexed; Carbo, carboplatin; m, months; pts, patients.

Radiological evaluation and immune-related unconventional pattern of response

In the previous Phase I–III trials, the efficacy of nivolumab has been evaluated using the Response Evaluation Criteria in Solid Tumors v1.1 (RECIST v1.1) guidelines [105], which are currently considered the gold standard. Nevertheless, the kinetics of response of new immunomodulatory compounds may differ from that of chemotherapy. Due to their mechanism of action, checkpoint inhibitors such as nivolumab can lead to tumor infiltration by activated T-cells, which can sometimes radiologically appear as an increased tumor burden soon after the start of treatment[106]. This "pseudo-progression" can eventually be followed by tumor response, in a time frame ranging from 6 weeks to 6 months. Since radiological features are not currently available to definitively discriminate pseudo-progression from real tumor spread, treatment continuation beyond radiological progression can be considered for clinically stable patients[106]. Moreover, taking into account these particular features, new guidelines for response evaluation of immune therapy have been proposed [106].

An unconventional pattern of response, described as RECIST v1.1-defined progressive disease followed by PR or SD as defined per protocol, was described in a relatively low percentage of NSCLC patients treated with nivolumab. Across the Phase II and III trials, ~3%–7% of patients experienced an unconventional response. In particular, out of 117 highly pretreated patients who received nivolumab in the Phase II trial [2, 37], 22 patients were treated beyond progression, and four (3.4% of the total) met criteria for unconventional benefit. This proportion was slightly higher in the two Phase III trials: an unconventional response was seen in nine (6.8%) out of 131 nivolumab-treated squamous NSCLCs (28 patients treated beyond progression)[3] and in 16 (5.5%) out of 292 nivolumab-treated non-squamous NSCLCs (71 patients treated beyond progression) [5].

Predictive value of PD-L1 expression and emerging predictors of response to anti-PD-1 therapy

In advanced NSCLCs, nivolumab monotherapy achieves RRs of ~20%[3, 5]. Therefore, predictive factors are desirable both to select patients who can more likely benefit from anti-PD-1 treatment and for economic reasons.

Various trials, evaluating different PD-1/PD-L1 pathway inhibitors in different tumor types, have described conflicting results about the role of PD-L1 expression on tumor cells in predicting the response to treatment. Therefore, a large meta-analysis (20 trials; 1,475 patients) was conducted in order to explore the role of PD-L1 as predictive factor [55]. Among the overall population (including patients with melanoma, NSCLC, and genitourinary cancer), treated with either a PD-1 or a PD-L1 inhibitor, a significantly higher RR was described for PD-L1-positive patients, compared to the PD-L1-negative patients (ORR: 34.1% vs 19.9%, P<0.0001). The difference was also significant in the subgroup of patients treated with nivolumab (absolute difference: 16.4%, 95% CI 10.0–22.7, P<0.0001) and among patients with NSCLC (absolute difference: 8.7%, 95% CI 1.1–15.5, P=0.02) [107]. However, this study also pointed out that a non-negligible proportion of PD-L1-negative patients still respond to anti-PD-1 or anti-PD-L1 treatments.

For nivolumab-treated NSCLC patients, available data are still controversial so far (**Table 3**). Results from a non-randomized subset of 61 pretreatment specimens from 42 patients enrolled in the Phase I trial suggested a role for PD-L1 expression in predicting response to nivolumab, with 36% OR among PD-L1-positive patients and no OR among PD-L1-negative patients[7]. Out of the ten patients with NSCLC evaluable for PD-L1 expression in this preliminary analysis, five were PD-L1 positive, and one of them (with 10% positive tumor cells) achieved a PR with nivolumab 10 mg/kg. However, in a bigger cohort of NSCLCs, no association was seen between PD-L1 status and either ORR or OS[9]. Data from Phase III trials favored nivolumab in squamous NSCLC despite PD-L1 expression[3], while for non-squamous NSCLCs, PD-L1 expression seemed to be predictive of better nivolumab efficacy in terms of ORR, PFS, and OS [5]. Across different trials, ORs and longer DORs have been registered both in PD-L1-positive and PD-L1-negative NSCLCs, even if numerically higher among positive tumors [2, 36], and no differences have been described for different levels of PD-L1 expression (1%, 5%, or 10% positive tumor cells) [2, 3, 5, 37].

Table 3 - Correlation between PD-L1 expression and clinical response to nivolumab in NSCLC

Study	Evaluable specimens (pts n)	PD-L1 cut-off (% positive tumor cells)	Findings
Topalian 2012[7] *Phase I (melanoma, NSCLC, prostate, CRC, RCC)*	61 specimens 42 pts *(18 mel., 10 NSCLC, 7 CRC, 5 RCC, 2 prostate)*	≥ 5%	25 pos/42 pts → OR: 9 pts (36%) 17 neg/42 pts → OR: none *Data suggestive for a relationship between PD-L1 expression and OR*
Gettinger 2015 [9][40] *Phase I (prolonged FU NSCLC)*	68 pts	≥ 5%	33 pos/68 pts → ORR: 15%; median OS: 7.8 m (5.6 to 21.7) 35 neg/68 pts → ORR: 14%; median OS: 10.5 m (5.2 to 14.8) *No association between PD-L1 status and ORR or OS*
Rizvi 2015[2] *Phase II (SqNSCLC) CheckMate 063*	76 pts	≥ 1% ≥ 5% ≥ 10%	PD-L1 pos (≥1%) → ORR: 20% PD-L1 neg (<1%) → ORR: 13% PD-L1 pos (≥5%) → ORR: 24% PD-L1 neg (<5%) → ORR: 14% PD-L1 pos (≥ 10%) → ORR: 24% PD-L1 neg (<10%) → ORR: 14% *OR numerically higher in PD-L1 positive NSCLCs; No differences among different levels of PD-L1 expression*
Brahmer 2015[3] *Phase III (SqNSCLC) CheckMate 017*	225 pts (117 received Nivolumab)	≥ 1% ≥ 5% ≥ 10%	PD-L1 pos (≥1%) → ORR: 17% PD-L1 neg (<1%) → ORR: 17% PD-L1 pos (≥5%) → ORR: 21% PD-L1 neg (<5%) → ORR: 15% PD-L1 pos (≥10%) → ORR: 19% PD-L1 neg (<10%) → ORR: 16% *PD-L1 expression has no predictive or prognostic value; Nivolumab is more effective than docetaxel despite PD-L1 level*
Borghaei 2015 [5] *Phase III (Non-SqNSCLC) CheckMate 057*	455 pts (231 received Nivolumab)	≥ 1% ≥ 5% ≥ 10%	PD-L1 pos (≥1%) → ORR: 31% PD-L1 neg (<1%) → ORR: 9% PD-L1 pos (≥5%) → ORR: 36% PD-L1 neg (<5%) → ORR: 10% PD-L1 pos (≥10%) → ORR: 37% PD-L1 neg (<10%) → ORR: 11% *Strong predictive association between PD-L1 expression and outcome (ORR, PFS, OS) at all expression levels*
Rizvi WCLC 2015[103] *Phase I (NSCLC) CheckMate 012*	113 pts (Nivo + Ipi)	≥1%	PD-L1 pos (≥1%) → ORR: 8-48% (across different dose regimens[§]) PD-L1 neg (<1%) → ORR: 0-22% (across different dose regimens[§]) *Clinical activity was observed regardless of tumor PD-L1 expression; Preliminary evidence of greater activity in ≥1% PD-L1 positive tumors*

Note: *[a]Dose regimens included: nivolumab 1 mg/kg + ipilimumab 1 mg/kg, q3w; nivolumab 1 mg/kg, q2w + ipilimumab 1 mg/kg, q6w; nivolumab 3 mg/kg, q2w + ipilimumab 1 mg/kg, q12w; nivolumab 3 mg/kg, q2w + ipilimumab 1 mg/kg, q6w.*
Abbreviations: NSCLC, non-small-cell lung cancer; pts, patients; pos, positive; CRC, colorectal cancer; RCC, renal cell carcinoma; OR, objective response; neg, negative; FU, follow-up; ORR, overall response rate; PD-L1, programmed death 1 – ligand 1; PFS, progression-free survival; OS, overall survival.

The predictive role of PD-L1 expression has also been investigated in trials evaluating other anti-PD-1 compounds, such as pembrolizumab [51, 56]. In the Phase II/III study Keynote-010, PD-L1-positive (ie, PD-L1 expression ≥1% of tumor cells) NSCLC patients treated with pembrolizumab achieved a longer median OS compared to those receiving docetaxel (pembrolizumab 2 mg/kg: 10.4 months; pembrolizumab 10 mg/kg: 12.7 months; docetaxel 75 mg/m^2: 8.5 months). This survival advantage was higher for patients with ≥50% of PD-L1-positive tumor cells, despite the dose of pembrolizumab they received (HR 0.54 for pembrolizumab 2 mg/kg vs docetaxel, 95% CI 0.38–0.77, P=0.0002) [51].

Most nivolumab trials evaluated PD-L1 expression retrospectively on archival tumor samples, using an automated immunohistochemical assay (Dako Denmark A/S, Glostrup, Denmark). Different assays are currently available for the evaluation of PD-L1 expression, but an FDA blueprint project is ongoing to solve the differences between the registered kits[108, 109]. Unfortunately, the population tested is heterogeneous, and the PD-L1 expression in tumors seems to be heterogeneous [108]. This makes the interpretation of current data uncertain. However, the absence of univocal results suggests that PD-L1 expression might not be the only predictor of response to immune checkpoint inhibitors[109].

Recent data support the hypothesis that tumor's mutation burden could influence the response to PD-1 inhibitors. In fact, the efficacy of PD-1 inhibitors such as nivolumab is based on the ability of T-cells to recognize tumor-related antigens that are presented on the tumor cell surface by major histocompatibility complexes (**Figure 2**). In particular, neo-antigens (ie, neo-epitopes deriving from tumor-specific DNA mutations) seem to play an important role in tumor immune control[100], as suggested by the sporadic observation of systemic tumor response after local radiotherapy [110]. The likelihood of formation of neo-antigens that can be recognized by host T-cells is expected to be higher in tumors with a high mutational load, in particular if this is higher than ten somatic mutations per megabase pair (corresponding to 150 nonsynonymous mutations within expressed genes) [100].

Among different tumor types, there is high variability in mutation frequency, but differences can also be seen within the same tumor type[111, 112]. For NSCLCs, substantial differences have been described between smokers and never smokers both in terms of mutational burden and affected genes [113]. Smoking-induced lung cancers are characterized by a higher number of mutations per megabase pair compared to tumors of never smokers [112, 114, 115]. In particular, Govindan et al described a median of 10.5 mutations per megabase pair (range: 4.9–17.6) in smokers and a median of 0.6 (range: 0.6–0.9) in never smokers[116]. Recently, Rizvi et al have demonstrated a significantly improved efficacy of anti-PD-1 treatment for NSCLCs with a high nonsynonymous mutation burden, in terms of ORR, durable clinical benefit (ie, PR or SD lasting ≥6 months), and PFS [112]. Moreover, the benefit was greater for tumors harboring the "smoking signature" (ie, transversion-high [TH])[113] compared to those with transversion-low (TL) tumors (ORR: TH 56% vs TL 17%, P=0.03; durable clinical benefit: TH 77% vs TL 22%, P=0.004; PFS: TH not reached vs TL 3.5 months, P=0.0001) [112]. Lastly, recent evidence that tumors with mismatch-repair deficiency achieve higher ORR and survival compared to mismatch-repair-proficient ones seems to support the hypothesis of a role for tumor mutation load and neo-antigens in predicting the response to anti-PD-1 treatments [117, 118].

Safety and tolerability

In general, nivolumab is well tolerated (**Table 4**), and patients' performance status has been reported not to affect treatment tolerability[101]. In two Phase III trials, nivolumab was compared to docetaxel and was found to induce fewer grade 3–4 events than chemotherapy (7%–10% vs 54%–55%, respectively) [3, 5]. Across different trials [2, 3, 5, 9], treatment-related adverse events of any grade were reported in 58%–74% of the patients. The most frequent ones were fatigue, decreased appetite, and asthenia. Grade 3 or 4 adverse events were reported in 7%–17% of the patients, and the most common event was fatigue. No clear relationship between the occurrence of events and dose level or treatment duration was found [3, 9].

Table 4 - Most common nivolumab-related immune-mediated adverse events and reported frequency in the main clinical trials

Study	Pts (n)	Pneumonitis		Diarrhea		Hypothyroidism		Skin toxicity		Renal toxicity	
		Any grade	*Grade 3-4*	*Any grade*	*Grade 3-4*	*Any grade*	*Grade 3-4*	*Any grade*	*Grade 3-4*	*Any grade*	*Grade 3-4*
Gettinger 2015 [9] *Phase I (Prolonged FU NSCLC)* NCT00730639	129 NSCLC	8 (6%)	3 (2%)	13 (10%)	1 (1%)	N/R	N/R	*Rash*: 9 (7%)	None	N/R	N/R
Rizvi 2015 [2] *Phase II (SqNSCLC) CheckMate 063* NCT01721759	117 NSCLC	6 (5%)	4 (3%)	12 (10%)	3 (3%)	3 (3%)	None	*Rash*: 13 (11%)	1 (1%)	4 (3%)	None
Brahmer 2015 [3] *Phase III (SqNSCLC) CheckMate 017* NCT01642004	131 NSCLC Nivo	6 (5%)	1 (1%)	10 (8%)	None	5 (4%)	None	*Rash*: 5 (4%)	None	*Creatinine Increase* 4 (3%) *Nephritis* 1 (1%)	None 1 (1%)
Borghaei 2015[5] *Phase III (Non SqNSCLC) CheckMate 057* NCT01673867	287 NSCLC Nivo	8 (3%)	3 (1%)	22 (8%)	2 (1%)	19 (7%)	None	*Rash*: 27 (9%)	1 (<1%)	*Creatinine Increase* 5 (2%) *Renal failure: 1* (1%)	None None

Note: Data are presented as n (%).
Abbreviations: pts, patients; FU, follow-up; NSCLC, non-small-cell lung cancer; N/R, not reported.

irAEs were reported in approximately half of NSCLC patients treated with nivolumab across different trials. The most common irAEs were skin toxicity (5%–16%, consisting mainly in rash and pruritus), gastrointestinal events (8%–12%), and pneumonitis (3%–6%), and in most cases, they were of low grade.[2, 3, 9]. Other less common irAEs included endocrinopathies (4%–7%), elevation of blood liver function parameters (1%–3%), nephrotoxicity (2%–3%, mainly consisting in blood creatinine elevation), and rare infusion reactions (1%–3%)[2, 3, 5, 9].

Across different irAE categories, median time to onset (TTO) and time to resolution (TTR) ranged widely in the two Phase III trials by Brahmer et al (TTO: 0.3–17.6 weeks; TTR: 0.3–not reached)[3] and Borghaei et al (TTO: 0.1–31 weeks; TTR: 0.1–not reached)[5]. The

longest median TTO was registered for endocrine, hepatic, and pulmonary toxicities. Most of nivolumab-related adverse events were manageable with supportive care and glucocorticoids treatment, as per protocol.

Table 5 - Management of the most common irAEs

irAE	Grade 1	Grade 2	Grade 3-4
Pneumonitis	Consider discontinue treatment X-ray every 3 days If no improvement: Treat like grade 2	Discontinue treatment Start prednisone 1mg/kg/d, until resolved to grade 0-1 X-ray every 3 days If no improvement: Treat like grade 3-4	Discontinue treatment X-ray every 3 days Consider bronchoscopy/biopsy Start prednisone 1mg/kg/d, until resolved to baseline Taper prednisone in 6 weeks If no improvement (in 48 hours): Consider other immunosuppressive medication
Diarrhea	Continue treatment Start symptomatic treatment (ie loperamide)	Discontinue treatment Start symptomatic treatment (ie loperamide) Consider colonoscopy If no improvement: Treat like grade 3-4	Discontinue treatment Start prednisolone 1-2mg/kg/d until grade 0-1 Colonoscopy IV hydration and other (symptomatic) treatment of grade 3-4 diarrhea Taper prednisone in 5 weeks If no improvement: Consider infliximab
Hypothyroidism	Continue treatment Consider substitution therapy	Continue treatment Consider substitution therapy	MRI hypophyses Exclude other hormonal dysfunction Consult endocrinologist If abnormalities: Discontinue treatment Start prednisolone 1-2mg/kg/d
Skin toxicity	Continue treatment Consider local or oral treatment (ie topical steroids) If no improvement (in 2 weeks): Consider a biopsy and oral prednisone	Continue treatment Consider local or oral treatment (ie topical steroids) If no improvement (in 2 weeks): Consider a biopsy and oral prednisone	Discontinue treatment Consult dermatologist Start prednisone 1-2mg/kg/d, until resolved to grade 1 Taper prednisone in 5 weeks
Renal toxicity	Continue treatment	Discontinue treatment Check creatinine every 3 days Start prednisone 1mg/kg/d If no improvement (in 7 days): Treat like grade 3-4	Discontinue treatment Creatinine every 3 days Start prednisone 1mg/kg/d, until resolved to grade 1 Taper prednisone in 5 weeks

Note: Toxicity grading: as defined by CTCAE.
Abbreviations: irAEs, immune-related adverse events; IV, intravenous; MRI, magnetic resonance imaging; CTCAE, Common terminology criteria for adverse events.

The most common irAE leading to nivolumab discontinuation was pneumonitis [3, 9]. Grade 3–4 pneumonitis appeared in 1%–3% of the patients, and was generally manageable using corticosteroid treatment. In the trial of Borghaei et al[5], four patients (1%) experienced a grade 3–4 pulmonary adverse event (three pneumonitis; one interstitial lung disease). They were all treated with immune-modulating medication, and 75% of the events resolved completely. In the Phase II trial by Rizvi et al, all patients with pneumonitis were treated with steroids, and their median TTR was 3.4 weeks (1.6–13.4 weeks) [2]. Unresolved pneumonitis led to toxic death in three cases, all of them in the Phase I trial [9].

Diarrhea is another common irAE (8%–10%), sometimes associated with colitis [2, 3, 5, 9]. Therefore, as in the management of ipilimumab-related irAEs, with a persistent grade 2 diarrhea, a sigmoidoscopy or colonoscopy could be considered to rule out colitis [119]. Nevertheless, a grade 3–4 colitis was only reported in <1% of the patients overall. These patients improved after treatment with either supportive care or immunosuppressive therapy. When there is no improvement in 48–72 hours, infliximab could be an alternative [119].

Most commonly reported endocrine irAEs are thyroid impairments, such as hypothyroidism [2, 3, 5]. Hypophysitis has not been reported. TTR was not reached for endocrinopathies in both Phase III trials [3, 5], with a proportion of patients requiring prolonged substitution therapy with thyroid hormones. No grade 3 or 4 events were described in patients treated with nivolumab.

Treatment-related deaths were reported in two trials. In the Phase I trial by Gettinger et al [9], three cases of treatment-related deaths were described, associated with pneumonitis. Two of the patients had unresolved grade 4 pneumonitis, and the other one, grade 5. Rizvi et al [2] described two nivolumab-related deaths. One of the patients had rapid tumor progression and bronchial obstruction. An inflammatory component caused by nivolumab could not be ruled out because a bronchoscopy or autopsy was not performed. The second patient died of ischemic stroke 41 days after the only dose of nivolumab he got. Both these patients had multiple comorbidities.

In general, grade 1 or 2 irAEs are treated symptomatically (eg, loperamide for diarrhea), and discontinuation is not always necessary. For grade 3 and 4 irAEs, the treatment with nivolumab should be discontinued, and steroids (or other immunosuppressive therapy) should be started. For symptomatic endocrinopathy, substitution therapy might be required (Table 5) [119].

Other anti-PD1 and anti-PD-L1 compounds, such as pembrolizumab, showed a comparable safety pattern. No trials are currently available comparing the safety of these compounds [39].

Patient-focused perspectives: QoL and patient-reported outcomes

Given the peculiar spectrum of immune-related side effects among nivolumab-treated patients, the evaluation of their QoL is as relevant as the drug's clinical activity to make a comprehensive comparison with standard treatments. Few data are currently available, which suggest a good QoL for patients treated with nivolumab. During the 2015 ASCO meeting, patient-reported outcomes from subjects with advanced melanoma,

treated with either nivolumab or dacarbazine in CheckMate 066 trial, were reported. QoL questionnaire completion rates were 70% in the nivolumab arm and 64.9% in the dacarbazine arm. No improvement from basal QoL was described for dacarbazine-treated patients. On the contrary, nivolumab-induced QoL improvements from week 7 to week 61, registered with EuroQoL-Five Dimension questionnaire (EQ-5D), utilities and visual analog scale (VAS) scores [120]. Similarly, in CheckMate 067 trial, nivolumab led to early QoL improvements compared to ipilimumab [121]. Initial data are also available for nivolumab–ipilimumab combination regimens. These show that quality of life can be maintained at a similar level as with ipilimumab alone [121, 122].

For NSCLC, the only data available so far come from CheckMate 017 trial. In this study, the QoL questionnaire completion rates were 71.9% (97/135 patients) for the nivolumab arm and 64.2% (88/137 patients) for the docetaxel control group. A significant and progressive improvement in QoL (EQ-5D and EQ-VAS scores) was observed for subjects receiving nivolumab during the first year of treatment. EQ-VAS score was statistically higher than baseline at weeks 12, 20, 36, and 48 ($P \leq 0.05$), and similar results have been observed with EQ-5D index. Conversely, QoL for patients in the docetaxel arm showed no differences from baseline during their shorter treatment period [123]. Results from CheckMate 057 trial are still awaited.

Conclusion and future perspectives

In the recent years, new immune-modulating agents have emerged as effective treatments for the management of different tumors. In particular, nivolumab has been demonstrated to achieve a survival improvement over chemotherapy in patients with advanced NSCLC[3, 5] with a fraction of long-term survivors and a manageable toxicity profile. Given these striking results, nivolumab has recently been approved in the US and in Europe as second-line monotherapy for metastatic NSCLCs, of both squamous and non-squamous histologies. However, many questions are still open. Patients' selection is currently one of the biggest issues, both for treatment optimization and economic reasons. PD-L1 expression by tumor cells seems not to be sufficient to discriminate responders versus nonresponders, and new predictive factors are now under investigation. Tumor's mutational burden and neo-antigens are emerging as promising predictive factors [100, 112] and new diagnostic techniques are emerging to allow fast DNA sequencing, such as next-generation sequencing[124]. However, their applicability in clinical practice still has to be defined together with conclusive data of ongoing trials. The role of nivolumab in the treatment of NSCLC in other clinical settings still has to be defined. A number of ongoing trials are currently investigating its efficacy as first-line and adjuvant therapy. The use of nivolumab in combination with other systemic agents is promising, in particular when combined with other immune checkpoint inhibitors. Finally, the duration of administration of the checkpoint inhibitors is not yet defined. Studies addressing this issue are ongoing. Immunotherapy is opening new perspectives for the treatment of lung cancer, giving new effective options for this highly fatal disease, and new results from the ongoing trials are awaited in the next years.

2 Pembrolizumab for the treatment of non-small cell lung cancer

M. Muller*, R.D. Schouten*, C.J. De Gooijer, P. Baas.

Expert review of anticancer treatment, 2017. 17(5): 399-409

Abstract

Introduction: In the last years, a spectacular development of immunotherapeutic agents aimed at the PD-1/PD-L1 axis has taken place. This development of these checkpoint inhibitors has greatly influenced our approach to the treatment of lung cancer in first and second line. The limited toxicity profile and the ability to treat for prolonged periods, even in smokers, is a welcome expansion of the therapeutic arsenal of the oncologist.
Areas covered: This review highlights the results of recent clinical trials on pembrolizumab for the treatment of non-small cell lung cancer. The authors discuss both first and second line treatment with pembrolizumab as monotherapy and in combination therapies. Additionally, implications of the PD-L1 immunohistochemistry assay with the 22C3 antibody and its use in clinical practice and trials is discussed.
Expert commentary: A higher overall response, overall survival and a moderate toxicity profile is observed with the use of pembrolizumab, compared to chemotherapy, in both first and second line. These promising results have already translated into the registration of pembrolizumab in first and second line in patients with a high expression of PD-L1. However, as PD-L1 staining does not sufficiently discriminate responders from non-responders for all checkpoint inhibitors, there still is a need for a better predictive biomarker.

Key issues

- In the setting of advanced or metastasized non-small cell lung cancer, pembrolizumab is FDA approved for:
    - The treatment of patients who are progressive on or after platinum based chemotherapy and are at least 1% PD-L1 positive.
    - First line treatment of patients whose tumor is ≥50% PD- L1 positive, using the validated IHC assay with the 22C3 antibody.
- Pembrolizumab compared to docetaxel showed a doubling of overall response rate and 24 months overall survival.
- PD-L1 positivity (TPS ≥50%) is associated with higher ORR, PFS and OS for the use of pembrolizumab.
- PD-L1 expression may be highly variable between different tumor sites, leading to a possible false negative result.
- The toxicity profile of pembrolizumab is superior over chemotherapy in both first and second line.
- Pseudoprogression can complicate a proper response evaluation, but is an uncommon event in lung cancer.

1. Introduction

Chemotherapy has been the standard treatment for patients with disseminated non-small cell lung cancer (NSCLC) for many years. In the last two decades, several novel therapeutic agents have been developed [13-15]. Major progress has been made in the field of targeted therapy. These agents have become part of the regular treatment plan for a subgroup of patients. The efficacy of these drugs depends heavily on the presence of oncogenic driver mutations, for example aberrant EGFR or ALK. However, only a minority of patients are candidates for this kind of treatment, underscoring the need for new treatment avenues to be explored [14, 17, 116].

The unraveling of the interaction between various proteins with functions in the immunologic response in the tumor microenvironment, initiated the development of immunotherapeutic drugs aimed at these proteins. In particular, the interaction between programmed cell death 1 (PD-1) and programmed death ligand-1 (PD-L1, also known as B7 homolog1, or B7-H1) was of interest. This interaction (the PD pathway) was shown to be a potential immune escape mechanism for tumor cells [19, 98, 125]. In the last decade, a novel category of therapeutic agents, so called checkpoint inhibitors, has emerged. These immunotherapeutic drugs have provided new opportunities in treating a variety of cancers including lung cancer. Several of these agents have been developed and tested in clinical trials and some have resulted in a significant improvement in clinical outcome [3, 5, 13, 15, 18, 24, 39, 40, 42, 104, 126-128].

In this review, we provide a brief overview of immunotherapeutic drugs aimed at the PD pathway, which are currently available in the clinic or in various testing phases in clinical trials. We focus on the development and characteristics of pembrolizumab, one such checkpoint inhibitor, discussing its current uses and future perspectives in treatment of NSCLC.

Pembrolizumab (Keytruda®, lambrolizumab, MK-3475) is a humanized monoclonal antibody (class IgG4/kappa) which was granted accelerated approval for treatment of patients with metastatic NSCLC, whose tumors show expression of PD-L1 [41, 129]. Pembrolizumab binds with high affinity to the PD-1 receptor, blocking its association with PD-L1 and PD-L2. This prevents an inhibitory signal from cancer cells to cytotoxic CD8 T cell [92, 130, 131] (Figure 1). It is the second anti-PD-1 antibody to be approved for second-line treatment of metastatic NSCLC, and the first to be approved for first-line treatment, provided it is accompanied by a diagnostic test selecting a specific subgroup of patients [15, 41, 132]. The goal of this article is to review the available clinical data on treatment of NSCLC with pembrolizumab, discussing efficacy, safety, quality of life (QoL) and the role of PD-L1 testing.

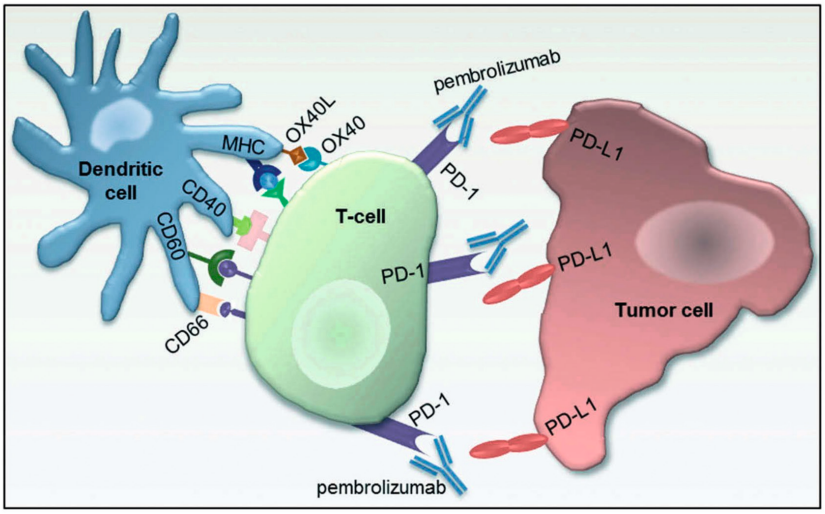


Figure 1 - Schematic of ligand-receptor pairs involved in inhibition and stimulation of T-cell antitumor activity.
Pembrolizumab blocks the PD-1/PD-L1 interaction, preventing immune escape by tumor cells.

1.1 Current needs, available therapies, competitor compounds

The class of checkpoint inhibitors receiving most attention and achieving best results in treatment of NSCLC in recent years are monoclonal antibodies (mAb) blocking the PD-1/PD- L1 (and PD-1/PD-L2) interaction. These can be divided into two groups: anti-PD-1 or anti-PD-L1 mAb.

The first anti-PD-1 mAb receiving approval for second-line treatment of NSCLC was nivolumab (Opdivo®, MDX-1106, BMS-936558, ONO-4538), showing objective responses in 17% of patients and an increased median overall survival (OS) of 14.9 months across all NSCLC patients at a dose of 3 mg/kg administered every other week [3, 5, 9, 132]. Pembrolizumab, an anti-PD-1 mAb, is registered for treatment of NSCLC in both first line (dependent on PD-L1 expression, discussed later in this article) and second line, and has shown increasing efficacy in the presence of PD-L1 expression on tumor cells [39, 40, 42, 127, 128, 133].

Other monoclonal antibodies aimed at PD-L1 have been developed as well. Atezolizumab (Tecentriq®, MPDL3280A) has recently been approved by the US FDA for treatment of NSCLC. Compared to docetaxel, it showed a significant improvement in OS of 4.2 months (13.8 months versus docetaxel, OS 9.6 months) in patients with previously treated NSCLC. This result was independent of both the histological subtype and level of PD-L1 expression. OS was longer in patient groups with higher levels of PD-L1 expression as was observed in the studies with pembrolizumab [104, 107, 134, 135]. The anti-PD-L1 mAb durvalumab (MEDI4736) is currently tested in phase III trials in multiple settings, including first, second and later lines of treatment, multiple stages of NSCLC, and in multiple combinations, including adjuvant after intended curation with chemoradiation therapy. Results of these trials are yet to be published [136-138]. Avelumab (MSB0010718C), a third anti-PD- L1 mAb tested in the large phase I 'Javelin'

trial, showed an objective response rate (ORR) of 22.4% of NSCLC patients not selected by histologic subtype of PD-L1 expression in first-line setting. The 24-week progression-free survival (PFS) rate was 37.2%. Further results are pending [139]. Last, the anti-PD-L1 mAb codenamed BMS-936559 has been tested in several solid tumors in a phase I trial and an objective response was seen in 5 out of 49 patients with NSCLC. This agent is not being further developed at this time [140, 141].

2. Pharmacology of pembrolizumab

2.1 Pharmacodynamics

Pembrolizumab is a highly selective antibody that blocks binding of PD-1 to PD-L1 and PD-L2 [142]. PD-1 was discovered in 1992 as member of the B7-CD28 superfamily. The immune regulatory role of PD-1 was shown in PD-1-deficient mice by inducing peripheral tolerance and antigen-specific immunity in infection and cancer. PD-1 is expressed on activated T cells (both CD8+ and CD4+), natural killer (NK) cells, antigen presenting cells (APC), monocytes, and B cells. Inflammatory cytokines produced by these cells induce expression of PD-L1 and PD-L2 on APCs. In this negative feedback loop, PD-1/PD-L1 and PD-1/PD-L2 interactions induce inhibitory signaling to T- and B cells, leading to decreased cytokine production and decreased antibody formation, and consequently to inhibition of antitumor immunity [143]. Thus, blocking interactions of the PD-1 receptor by pembrolizumab prevents inhibition of the immune response by tumor cells.

2.2 Pharmacokinetics

Ahamadi et al. pooled study data of 2195 patients who were treated for various types of cancer with pembrolizumab at intravenous doses of 1–10 mg per kilogram bodyweight (mg/kg) and in intervals of two or three weeks. Data were collected from three trials: KEYNOTE-001, KEYNOTE-002, and KEYNOTE-006 [144]. The pharmacokinetics of pembrolizumab were similar to other monoclonal antibodies. The clearance of pembrolizumab was low with 0.22 L/day. Central volume of distribution (V_c) was limited at 3.48 l and interindividual variability was low to moderate, with 38% for clearance and 21% for V_c. The elimination half-life was estimated at 14–27 days, with an apparently linear serum exposure with steady-state dosing in the range of 0.1–10 mg/kg [40, 133]. Steady state was reached after approximately 129 days and a modest accumulation of about 2.2-fold, based on AUC calculation at a three- weekly dosing schedule [144]. Effects of sex, baseline Eastern Cooperative Oncology Group Performance Status (ECOG-PS), ipilimumab treatment prior to pembrolizumab exposure, renal function, hepatic function, and tumor type are lacking in clinical significance [9, 144]. Recently presented data show that a flat dose of 200 mg results in equal serum exposure levels as compared to a three-weekly dose of 2 mg/kg, as tested in 152 individuals [145].

3. Clinical efficacy of pembrolizumab

3.1 Pembrolizumab monotherapy

The first study with pembrolizumab for use in treatment of metastatic NSCLC, the KEYNOTE-001 [40], was published in April 2015. In this phase I trial, with both pretreated and non-pretreated patients, 495 patients received at least one cycle of pembrolizumab,

at a dose of 2 mg/kg once every 3 weeks (Q3W, *n* = 6), 10 mg/kg Q3W (*n* = 287) or 10 mg/kg Q2W (*n* = 202). The ORR in all patients in this trial was 19.4% (95% CI:16.0–23.3%), PFS 3.7 months (median, range: 2.9–4.1 months) and OS 12.0 months (median, range: 9.3–14.7 months) (Table 1).

Table 1 - Overview of trials evaluation pembrolizumab in NSCLC

Name of study NCT number	**Phase Line**	**Disease setting Treatment**	**Patients (N)**	**ORR (percentage (95% CI)**	**PFS (months (95% CI))**	**OS (months)**
KEYNOTE 001 [40] NCT01295827	Phase I ≥ *1st line*	Treated and untreated A) Pembrolizumab 2 mg/kg Q3W B) Pembrolizumab 10 mg/kg Q3W C) Pembrolizumab 10 mg/kg Q2W	495 6 287 202	19.4% (16.0-23.2%)	3.7 months (2.9-4.1)	12.0 months (9.3-14.7)
KEYNOTE 001 [133] NCT01295827	Phase I ≥ *1st line*	Treated and untreated A) Pembrolizumab 2 mg/kg Q3W	52	15% (7-28%)	NA	NA
KEYNOTE 010 [39] NCT01905657	Phase II-III ≥ *2nd line*	PD-L1 positive ≥1% A) Pembrolizumab 2 mg/kg Q3W B) Pembrolizumab 10 mg/kg Q3W, C) Docetaxel 75 mg/m2 Q3W	1034 345 346 343	18% 18% 9% **P-values** A-C: P=0.0005 B-C: P=0.0002	3.9 months (3.1-4.1) 4.0 months (2.7-4.3) 4.0 months (3.1-4.2) **Hazard ratio** A-C: 0.88 (0.74-1.05, p=0.07) B-C: 0.79 (0.66-0.94, p=0.004)	10.4 (9.4-11.9) 12.7 (10.0-17.3) 8.5 (7.5-9.8) **Hazard Ratio** A-C: 0.71 (0.58-0.88, p=0.0008) B-C: 0.61 (0.49-0.75; p<0.0001)
KEYNOTE 021 [135] NCT02039674	Phase II *1st line*	Multicohort study A) Pembrolizumab + Chemotherapy B) Chemotherapy	123 59 62	55% (42-68%) 29% (18-41%)	13.0months (8.3-NR) 8.9 months (4.4-10.3) **Hazard ratio** 0.53 (0.31-0.91; p=0.010)	NR NR **Hazard ratio** 0.9 (0.42 – 1.91, P=0.39)
KEYNOTE 021 [146] Cohort A+B+C NCT02039674	Phase I ≥ *3rd line*	NSCLC Pembrolizumab + A) Carboplatin + Paclitaxel B) Carboplatin + Paclitaxel + Bevacizumab C) Carboplatin + Pemetrexed	74 25 25 24	57% (45-68%) 52% (31-72%) 48% (28-69%) 71% (49-87%)	10 months (6-NR) 10 (4-NR) NR (4.1-NR) 10 (6-15)	NR (17-NR) NR (11-NR) NR (NR-NR) NR (14-NR)
KEYNOTE 021 [147] Cohort D+H NCT02039674	Phase I ≥ *2nd line*	NSCLC A) Pembrolizumab + Ipilimumab	45	24% (13-40%)	6 months (1-17)	17 months (6-17)
KEYNOTE 024 [42] NCT02142738	Phase III *1st line*	PD-L1 Positive (≥50%) A) Pembrolizumab 200 mg Q3W B) Platinum-based chemotherapy	305 154 151	44.8% (36.8-53.0%) 27.8% (20.8-35.7%)	62.1% (53.8-69.4) 50.3% (41.9-58.2) **Hazard ratio:** 0.5 (0.37-0.68, P<0.001)	NR NR **Hazard ratio:** 0.6 (0.41-0.89; p=0.005)

Abbreviations: n: number, ORR: objective response rate, PFS: progression free survival, OS: Overall Survival, Q2W: one course every two weeks, Q3W: One course every three weeks. NR: Not reached, NCT: Number at clinicaltrials.gov.

Treatment naive patients had a slightly higher ORR of 24.8% (95% CI: 16.7–34.3%) compared to pre-treated patients: 18.0% (14.4–22.2%). In addition, a significant difference between smokers (ORR 22.5% (18.3–27.1%)) and nonsmokers (ORR 10.3% (5.6–17.0%)) was found. As a result of another trial in melanoma with pembrolizumab, which showed no differences between 2 mg/kg and 10 mg/kg [148], an extra cohort of 55 patients was added who received 2 mg/kg Q3 W pembrolizumab. The ORR in this group of patients was 15% (95% CI: 7–28%). In October 2015, based on this trial, the FDA approved pembrolizumab for the treatment of patients with metastatic NSCLC who are progressive on or after platinum based chemotherapy.

The KEYNOTE-010 trial [39] studies the effect of two doses of pembrolizumab in pretreated patients with NSCLC who expressed PD-L1. It was a randomized controlled phase II/III trial, with a total of 1034 patients, equally randomized to either pembrolizumab (2 mg/kg Q3W or 10 mg/kg Q3W) or docetaxel (75 mg/m2 Q3W). The ORR was 18% in both pembrolizumab groups compared to 9% in the docetaxel group, which was significant in both cases. Furthermore, the PFS was significantly different between pembrolizumab 10 mg/kg and docetaxel (Hazard Ratio (HR) 0.79 (95% CI: 0.66–0.94, $p = 0.004$)), but not between pembrolizumab 2 mg/kg and docetaxel (HR 0.88 (0.74–1.05, $p = 0.07$)) for patients who had tumors with only ≥1% of cells expressing PD-L1. The updated OS data of this trial and duration of clinical benefit of patients who completed their treatment with pembrolizumab (total treatment time of 24 months), were presented at the World Conference on Lung Cancer (WCLC) in Vienna in 2016 [149]. Median follow-up was 2.1 years (median, range: 1.5–3.0 years). Median OS was 10.5 months (range: 9.6–12.4) in the 2 mg/kg and 13.4 months (median, range: 11.2–17.0) in the 10 mg/kg group. Compared to docetaxel (median 8.6 months, range: 7.0–9.8) this was significantly longer, with Hazard Ratios of 0.72 (95% CI: 0.60–0.86, $p = 0.00017$) and 0.60 (95% CI: 0.49–0.72, $p < 0.0001$), respectively. The 24 months OS was 30.1% (95% CI: 25.0–35.4%) and 37.5% (95% CI: 32.2–42.9%) in both pembrolizumab groups, significantly higher than in the docetaxel group (14.5% (95 CI: 10.5–19.2%)).

In the first-line setting, pembrolizumab monotherapy was compared to standard chemotherapy in the KEYNOTE-024 phase III trial [42]. Only patients with tumors expressing ≥50% PD-L1 were included. The treatment schedule was a three weekly flat dose of 200 mg pembrolizumab. Results were first published in October 2016. Of the 305 patients assigned to this trial, 154 received pembrolizumab and 151 received chemotherapy. The ORR for pembrolizumab was 44.8% (95% CI: 36.8–53.0%) compared to 27.8% (95% CI: 20.8–35.7%) for chemotherapy. The median duration of response was not reached (NR) for the pembrolizumab group, compared to 6.3 months (range: 2.1–12.6 months) for the chemotherapy group. Median PFS was 10.3 months (range: 6.7-NR) compared to a median 6.0 months (range: 4.2–6.2) in the chemotherapy arm. The HR for the occurrence of a PFS event was 0.5 (95% CI: 0.37–0.68). OS was not reached in both cohorts.

3.2 Combination therapy

The combination of pembrolizumab with chemotherapy was examined in a small phase II trial [135]. This phase II trial, KEYNOTE-021, was part of a multicohort trial and compared the combination of pembrolizumab and chemotherapy ($n = 59$) to chemotherapy alone

(*n* = 62) in the first line. Response evaluation was done using RECIST version 1.1 assessed by blinded, independent central review. The ORR for combination therapy was 55% (95% CI: 42–68%) compared to 29% (95% CI: 18–41%). The benefit in the HR of PFS 0.53 (95% CI: 0.3–0.91; *p* = 0.010) was observed in favor of the combination therapy, with a median PFS in the combination group of 13.0 months (range: 8.3-NR), compared to a median 8.9 months (range: 4.4–10.3) in the chemotherapy alone group. For both groups, the median OS was not reached.

Data of five cohorts in the phase I part of the KEYNOTE-021 trial are not published yet. However, some preliminary data were presented at the annual meeting of the American Society of Clinical Oncology (ASCO) in 2016. Cohorts A, B, and C compared three combination therapies: pembrolizumab with (A) carboplatin + paclitaxel (*n* = 25), (B) carboplatin, paclitaxel and bevacizumab (*n* = 25), or (C) carboplatin + pemetrexed (*n* = 24). ORR for all patients was 57% (95% CI: 45–68%). So far, the combination with carboplatin and pemetrexed seems the most promising with an ORR of 71% (49–87%). Both PFS and OS data are incomplete and results are expected in June 2019 [146]. Cohorts D and H explore the combination therapy of pembrolizumab together with ipilimumab. The results of 45 enrolled patients showed an ORR of 24% (95% CI: 13–40%), median PFS of 6 months (range: 1–17), and median OS of 17 months (range: 6–17 months) [147].

KEYNOTE-189 and KEYNOTE-407 evaluated the combination of chemotherapy and pembrolizumab in first line in two randomized, placebo controlled trials. Results are expected in March 2019 (Figure 2). An overview of the most important other ongoing trials are listed in Table 2, including one trial exploring pembrolizumab after adjuvant therapy (KEYNOTE-091).

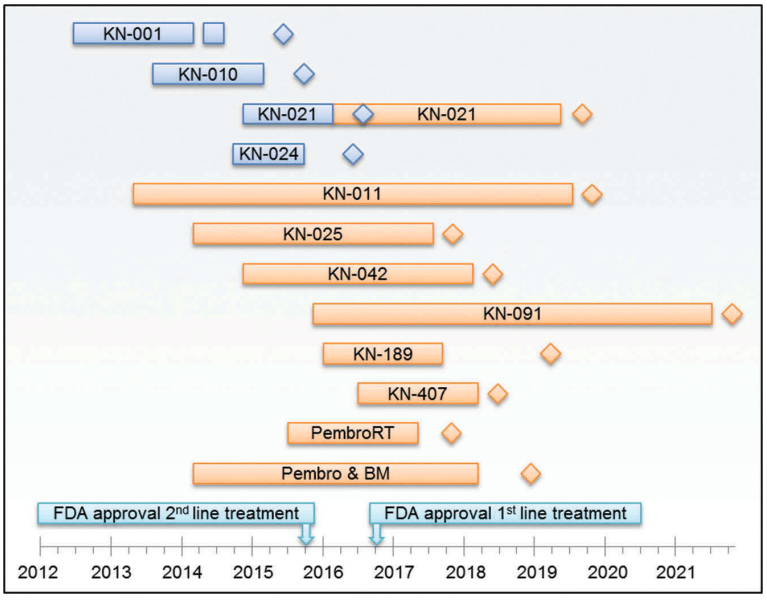


Figure 2 - Overview of published (blue) and current (orange) trials.
Diamonds represent (expected) results. Trials are corresponding with Table 1 (published trials) and 2 (ongoing trials).

Table 2 - Most important ongoing trials with pembrolizumab

Name of study NCT number	Phase Line	Disease setting Treatment	Estimated enrollment of patients (n)
KEYNOTE 011 NCT01840579	Phase I *2nd line*	Cohort B + C: Confirmed NSCLC B) Pembrolizumab + carboplatin + pemetrexed C) Pembrolizumab + carboplatin + (nab-) paclitaxel	48
KEYNOTE 025 NCT02007070	Phase I *2nd line*	PD-L1 positive (≥1%) A) Pembrolizumab 10 mg/kg Q3W	38
KEYNOTE 042 NCT02220894	Phase III *1st line*	PD-L1 positive A) Pembrolizumab 200 mg Q3W B) Chemotherapy (investigator's choice)	1240
KEYNOTE 091 (PEARLS) NCT02504372	Phase III *Post adjuvant treatment*	Stage IB, II - IIIA A) Pembrolizumab 200 mg Q3W B) Placebo	570
KEYNOTE 189 NCT02578680	Phase III *1st line*	Non-Squamous A) Pembrolizumab + chemotherapy B) Placebo (saline) + chemotherapy	570
KEYNOTE 407 NCT02775435	Phase III *1st line*	Squamous A) Pembrolizumab + chemotherapy B) Placebo (saline) + chemotherapy	560
Pembrolizumab and brain metastases NCT02085070	Phase II	Brain metastases + NSCLC A) Pembrolizumab	64
PEMBRO-RT NCT02492568	Phase II *2nd line*	Advanced NSCLC A) Pembrolizumab after SBRT B) Pembrolizumab alone	74

Abbreviations: *NCT: Number at clinicaltrials.gov. N: Number, Q2W: one course every two weeks, Q3W: one course every three weeks.*

3.3 PD-L1 expression and efficacy.

In the KEYNOTE-001 study, PD-L1 expression was assessed in tumor samples using two assays. First, the prototype assay which was developed and validated for use in a single accredited laboratory. Second, the clinical trial assay which was developed by DAKO, sponsored by MSD. Both assays are similar and the latter is now commercially available in an investigational kit. This immunohistochemistry (IHC) assay with anti-PD-L1 antibody clone 22C3 was tested by pathologists at DAKO and LabCorp facilities for accuracy and intra- pathologist reproducibility. The cut point was defined after a receiver operating characteristic (ROC) analysis, comparing multiple IHC scoring methods related to investigator-assessed immune-related response criteria (irRC) [40, 106]. As no major difference between scoring methods was observed, an operating point on the curve representing (simple) proportion score (PS, the proportion of viable cells expressing PD-L1 on their membrane on any level of intensity) was selected, which correlated with a Youden Index of 0.45–0.50. This operating point corresponded to a cutoff value of 50% (PS). The predictive value of the assay did not improve by including PD-L1 expression on T cells [40].

In the clinical setting, PD-L1 assessment is performed by scoring the percentage of viable tumor cells showing membranous staining of PD-L1 (tumor proportion score, or TPS). In the KEYNOTE-001 study, PD-L1 positivity was defined as a TPS of ≥1%. For defining the PD-L1 cut-of a training cohort was formed with 182 patients, of whom 129 had measurable disease according to RECIST 1.1 criteria.

Table 3 - ORR, PFS and OS on pembrolizumab, stratified by PD-L1 expression

Name of study NCT number	Treatment	Cut-Off	Patients (n)	ORR (percentage (95% CI))	PFS in months (median (range))	OS in months (median (range))
KEYNOTE 001 [40] NCT01295827	Pembrolizumab All regimes: 1) 2 mg/kg Q3W 2) 10 mg/kg Q3W 3) 10 mg/kg Q2W	Training: ≥50% 1-49% <1% Validation ≥50% 1-49% <1%	38 43 40 73 103 28	34.2% 9.3% 10.0% 45.2% 16.5% 10.7%	4.5 (1.9-12.5) 2.1 (2.0-2.9) 2.1 (1.8-2.5) 6.4 (4.2-NR) 4.1 (2.3-4.4) 4.0 (2.1-6.2)	13.7 (6.9-NR) 5.9 (4.2-8.2) 6.7 (3.9-10.0) NR (NR-NR) 10.6 (7.3-NR) 10.4 (5.8-NR)
KEYNOTE 001 [133] NCT01295827	Pembrolizumab: 1) 2 mg/kg Q3W 2) 10 mg/kg Q3W 3) 10 mg/kg Q2W	≥50% 1-49% <1% ≥50% 1-49% <1% ≥50% 1-49% <1%	23 23 4 42 49 18 31 43 9	30%(13-53%) 0% (0-15%) 25% (<1-81%) 48 (32-64%) 14 (6-27%) 6 (<1-27%) 39%(22-58%) 14% (5-28%) 11% (<1-48%)		
KEYNOTE 010 [39] NCT01905657	Pembrolizumab 1) 2 mg/kg Q3W 2) 10 mg/kg Q3W	≥50% ≥50%	139 151	30% 29%	5.0 (4.0-6.5) 5.2 (4.1-8.1)	14.9 (10.4-NR) 17.3 (11.8-NR)
KEYNOTE 021 [135] NCT02039674	Pembrolizumab + chemotherapy	≥50% 1-49% <1%	20 19 22	80% (56-94%) 26% (9-51%) 57% (34-78%)		
KEYNOTE 021 [146] Cohort A+B+C NCT02039674	Pembrolizumab + chemotherapy	≥50% 1-49% <1%	25 26 22	54% (33-73%) 55% (32-76%) 62% (45-78%)	15 (6-15) 14 (6-NR) 6 (4-NR)	17 (15-NR) NR (14-NR) 11 (7-NR)
KEYNOTE 021 [147] Cohort D+H NCT02039674	Pembrolizumab + ipilimumab	≥50% 1-49% <1%	6 18 21	17% 33% 19%	1 (0.2-6) NR (1-NR) 6 (1-17)	2 (0.3-NR) NR (3-NR) 17 (6-17)
KEYNOTE 024 [42] NCT02142738	Pembrolizumab 200 mg Q3W	≥50%	154	44.8% (36.8-53.0)	10.3 (6.7-NR)	NR

Abbreviations: *n: number, ORR: objective response rate, PFS: progression free survival, OS: Overall Survival, Q2W: one course every two weeks, Q3W: One course every three weeks. NR: Not reached. NCT: Number at clinicaltrials.gov.*

After stratifying results for PD-L1 expression, the cut-off for maximum efficacy was set at TPS ≥50%. In this training cohort, 38 patients had TPS ≥50%, with an ORR of 34.2% (95% CI: 19.6–51.4%) according to RECIST 1.1 (Table 3). 313 patients were included in the validation cohort, in which 73 patients had TPS ≥50%. Here, ORR was 45.2% (95% CI: 33.5–57.3%), median PFS in this cohort was 6.4 months (range: 4.2–NR), and median OS was not reached (range: NR–NR). For this analysis, only patients treated with pembrolizumab (mostly 10 mg/kg Q2W or Q3W) were evaluated [40]. In the added cohort

of the KEYNOTE-001 study, 23 out of 55 patients receiving 2 mg/kg Q3W pembrolizumab had TPS ≥50%. Here, ORR was 30% (95% CI: 13–53%) [133].

In the KEYNOTE-010 study [39], only patients with PD-L1 expression of at least 1% were included. In this study, 2222 screened patients were evaluable for their PD-L1 status, of whom 747 were PD-L1 negative (TPS <1%). For further PD-L1 positivity scoring in this trial, the cut-off was again set at TPS ≥50%. The number of patients with TPS ≥50% was similar in the three groups, namely *n* = 139 patients in the 2 mg/kg group, *n* = 151 in the 10 mg/kg group and *n* = 152 in the docetaxel group. ORR was 30% vs. 29% vs. 8%, respectively, which was higher compared to the total group (ORR: 18%, 18%, and 9%, respectively). PFS in both pembrolizumab groups was significantly superior over the docetaxel group with HR 0.59 (95% CI: 0.44–0.78, $p = 0.0001$) and HR 0.59 (95% CI: 0.45–0.78, $p < 0.0001$), respectively. Hazard ratios for OS were significantly lower in the pembrolizumab groups compared to the docetaxel group as well: 0.54 (95% CI: 0.38–0.77, $p = 0.0002$) and 0.50 (95% CI: 0.36–0.70, $p < 0.0001$), with a median OS of 14.9 months (range: 10.4–NR) and 17.3 months in the pembrolizumab groups, versus 8.2 months in the docetaxel group. In this trial, analysis of the PD-L1 expression was performed on both archival and fresh tumor samples [39]. At the ASCO Annual Meeting 2016, a poster was presented summarizing a comparative study on archival versus fresh tumor samples [150]. 44% of the patients had archival material and 56% had fresh samples, with a median of 250 days between sample collection and PD-L1 assessment (range: 2–2510 days) and 11 days (range: 1–371 days), respectively. For both groups, OS and PFS were significantly better compared to docetaxel, independent of age of the tumor samples. These data suggest archival tissue is sufficient for assessment of PD-L1 status, potentially saving patients the risks associated with obtaining tumor biopsies.

In the recently published KEYNOTE-021 trial [135], the combination of pembrolizumab with chemotherapy is compared to chemotherapy alone. Patients were stratified based on their PD-L1 status, classified in PD-L1 negative (<1%), or PD-L1 positive (1–49% or ≥50%). Patients who had a TPS ≥50% (*n* = 20) had a higher ORR (80% (95% CI: 56–94%)), compared to the TPS 1–49% group (ORR 26%, *n* = 19) and PD-L1 negative tumors (ORR 57%, *n* = 22), and compared to chemotherapy alone (ORR 35%, *n* = 17).

Finally, in the KEYNOTE-024 trial [14], patients were only included if they had TPS ≥50%. The group which received pembrolizumab (*n* = 154) had an ORR of 44.8% (95% CI: 36.8–53.0%), compared to 27.8% for chemotherapy. Based on this trial, the FDA approved pembrolizumab for first-line treatment of patients with metastasized NSCLC who are ≥50% PD- L1 positive, using the validated IHC assay with the 22C3 antibody.

Most studies show a relationship between the efficacy of pembrolizumab and level of PD-L1 expression on tumor cells. Expression is assessed by a validated assay using the 22C3 antibody on tumor biopsies, which are typically small in lung cancer. There have been several reports mentioning a heterogeneity of PD-L1 expression between tumor and metastatic sites, both lymph nodes and (distant) metastases. The rate of discordance between resections and matched biopsies may be up to 48%, suggesting that scoring of PD-L1 TPS on a single biopsy may be unreliable. Further research on this subject is needed, but being aware of the possible variability in PD-L1 expression within the same tumor is imperative [151-153].

4. Safety and tolerability

4.1 Pembrolizumab as monotherapy

Pembrolizumab blocks the PD-1 checkpoint, which is involved in preventing autoimmunity. Blocking this pathway may induce autoimmunity with unregulated activation of T cells directed at normal tissue. Therefore, toxicity profiles differ from chemotherapy or targeted therapy by inducing immune-mediated advance events (imAEs) [154].

The KEYNOTE-001 trial reported modest toxicity to pembrolizumab and the agent seems well tolerated. Treatment related adverse events of any grade were reported in 70.9% of patients. Grade 3 or higher toxicities were reported in 9.5% of treated patients. Less severe treatment-related side effect were low-grade diarrhea, fatigue and skin reactions. Most reported imAEs were hypothyroidism (6.9%), pneumonitis (3.6%), and infusion-related reaction (3.0%). Particularly, pneumonitis is an imAE of relevance in patients with NSCLC, as it may be life-threatening. During KEYNOTE-001, one patient died from this serious immunotherapy-related toxicity [40]. Chatterjee et al. reported an additional cohort of KEYNOTE-001 (pembrolizumab 2 mg/kg Q3W). In this cohort of 55 patients, 26 (47%) experienced a treatment-related adverse event, five (9%) of these were grade 3 or above. Data from the KEYNOTE-001 trial showed that the dose of pembrolizumab was not correlated with hazard for the occurrence of imAEs. However, the duration of treatment was positively correlated with the probability of experiencing an immune mediated adverse event [133].

The KEYNOTE-010 trial reported a comparable safety profile. 81% of the patients in the docetaxel group experienced an adverse event, which was higher than both pembrolizumab groups (63% in the 2 mg/kg group and 66% in the 10 mg/kg group). As was shown previously in KEYNOTE-001, fatigue and diarrhea were frequently reported. Grade 3–5 adverse events were reported in 13% and 16% of patients in the 2 mg/kg and 10 mg/kg groups, respectively, and in 35% of patients in the docetaxel group. The incidence of imAEs was similar in both pembrolizumab groups: 20% in the 2 mg/kg group and 19% in the 10 mg/kg group. Of 682 treated patients, six (<1%) died of pembrolizumab-related toxicities. Three deaths were due to pneumonitis, two due to pneumonia, and one due to myocardial infarction [16]. In the KEYNOTE-024 trial, pembrolizumab was compared to chemotherapy and diarrhea was the most reported adverse event in both treatment groups (14.3% and 13.3%, respectively, Table 4). Pneumonitis was reported in 5.8% of patients treated with pembrolizumab [42].

Table 4 - Toxicities compared with other agents

Study	First line treatment						Second line treatment					
Agent	**Pembrolizumab**		**Chemotherapy**		**Nivolumab**		**Pembrolizumab**		**Docetaxel**		**Nivolumab**	
Dose	**200 mg flat dose**				**3 mg/kg**		**2 mg/kg**		**75 mg/m2**		**3 mg/kg**	
	KEYNOTE-024 [42]		**KEYNOTE-024 [42]**		**CHECKMATE-012 [36]**		**KEYNOTE-010 [39]**		**KEYNOTE-010 [39]**		**CHECKMATE-057 [5]**	
Patients (N)	**N=145**		**N=150**		**N=52**		**N=339**		**N=309**		**N=287**	
	Any grade	**Grade 3-5**	**Any grade**	**Grade 3-5**	**Any grade**	**Grade 3-5**	**Any grade**	**Grade 3-5**	**Any grade**	**Grade 3-5**	**Any grade**	**Grade 3-5**
Any adverse event	73.4%	26.6%	90.0%	53.3%	71%	19%	63%	13%	81%	35%	69%	10%
Pneumonitis	5.8%	2.6%	0.7%	0.7%	6%	2%	4.1%	1.8%	1%	0.3%	3%	1%
Diarrhea	14.3%	3.9%	13.3%	1.3%	12%	2%	7.1%	0.6%	18.1%	2.3%	8%	1%
Hypothyroidism	9.1%	None	1.3%	None	6%	None	7.4%	None	0.3%	None	7%	None
Skin toxicity (rash, any)	3.9%	3.9%	None	None	19%	4%	10,70%	0.6%	5.1%	None	13%	0.3%
Renal toxicity (creat. increase)	1.9%	None	10.0%	0.7%	NR	NR	1.8%	None	None	None	1.7%	None
Fatigue	10.4%	1.3%	28.7 %	3.3%	29%	None	13.6%	1.2%	24.6%	3.6%	3%	25%

Abbreviations: *N: number, mg/kg: milligrams per kilogram bodyweight, NR: Not Reported, creat: creatinine*

4.2 Pembrolizumab in combination therapy

In the KEYNOTE-021 study, chemotherapy plus pembrolizumab was compared to chemotherapy alone. Adverse events of any grade were reported in 90% and 93%, respectively. Grade 3–5 toxicities were reported in 39% and 26% of treated patients. The incidence of imAE was higher in the pembrolizumab combination arm (22% versus 11%, respectively). Pneumonitis of any grade was reported in 3% of cases in the combination cohort and in none of the patients treated with chemotherapy alone. Other treatment groups of the KEYNOTE-021 included cohort A (pembrolizumab 2 or 10 mg/kg Q3W + carboplatin AUC 6 + paclitaxel 200 mg/m^2) and cohort C (pembrolizumab 2 or 10 mg/kg Q3W + carboplatin AUC 5 + pemetrexed 500 mg/m^2). In these cohorts, the rate of grade 3 or 4 treatment-related adverse events was 15% in cohort A, and 38% in cohort C. Toxicities included reversible transaminase elevation (*n* = 3 in cohort C), anemia (*n* = 1 in A, *n* = 2 in C), rash (*n* = 1 in A, *n* = 1 in C), and colitis (*n* = 2 in C). No treatment-related deaths occurred [43].

4.3 Quality of Life

The toxicity profile and derived therefrom the quality of life showed good results, as presented by Dr. Brahmer at the WCLC in Vienna 2016 [155]. The incidence of any grade of adverse event was 73% for pembrolizumab, compared to 90% for docetaxel, with incidences of grade 3–5 events of 27% and 53%, respectively. QoL was assessed, at baseline, at cycles 1, 2, and 3, and every 9 weeks thereafter, with two different questionnaires. The general QoL (QLQ-C30) showed a difference between the least square means of 7.82 (95% CI: 2.85–12.79, $p = 0.002$) in favor of pembrolizumab. The deterioration of typical lung cancer symptoms was assessed with the QLQ-LC13. Patients receiving pembrolizumab reported less deterioration compared to docetaxel (30% versus 39%). The hazard ratio (0.66, 95% CI: 0.44–0.97, $p = 0.029$) for time to deterioration was in favor of pembrolizumab as well.

5. Regulatory affairs

In October 2015, the FDA approved pembrolizumab for the treatment of patients with metastatic NSCLC who are progressive on or after platinum-containing chemotherapy. This approval was based on the KEYNOTE-001 trial. In October 2016, the FDA expanded the approval of pembrolizumab to first-line treatment of patients with NSCLC strongly positive for PD-L1 (TPS ≥50%, as assessed by IHC using the 22C3 anti- body). This was based on the KEYNOTE-010 and KEYNOTE-024 trials. The EMA stated in December 2016: ‘Keytruda as mono- therapy is indicated for the first-line treatment of metastatic non-small cell lung carcinoma (NSCLC) in adults whose tumors express PD-L1 with a ≥ 50% tumor proportion score (TPS) with no EGFR or ALK-positive tumor mutations. Keytruda as mono- therapy is indicated for the treatment of locally advanced or metastatic NSCLC in adults whose tumors express PD-L1 with a≥1% TPS and who have received at least one prior che- motherapy regimen. Patients with EGFR or ALK-positive tumor mutations should also have received targeted therapy before receiving Keytruda’ [156].

6. Expert commentary

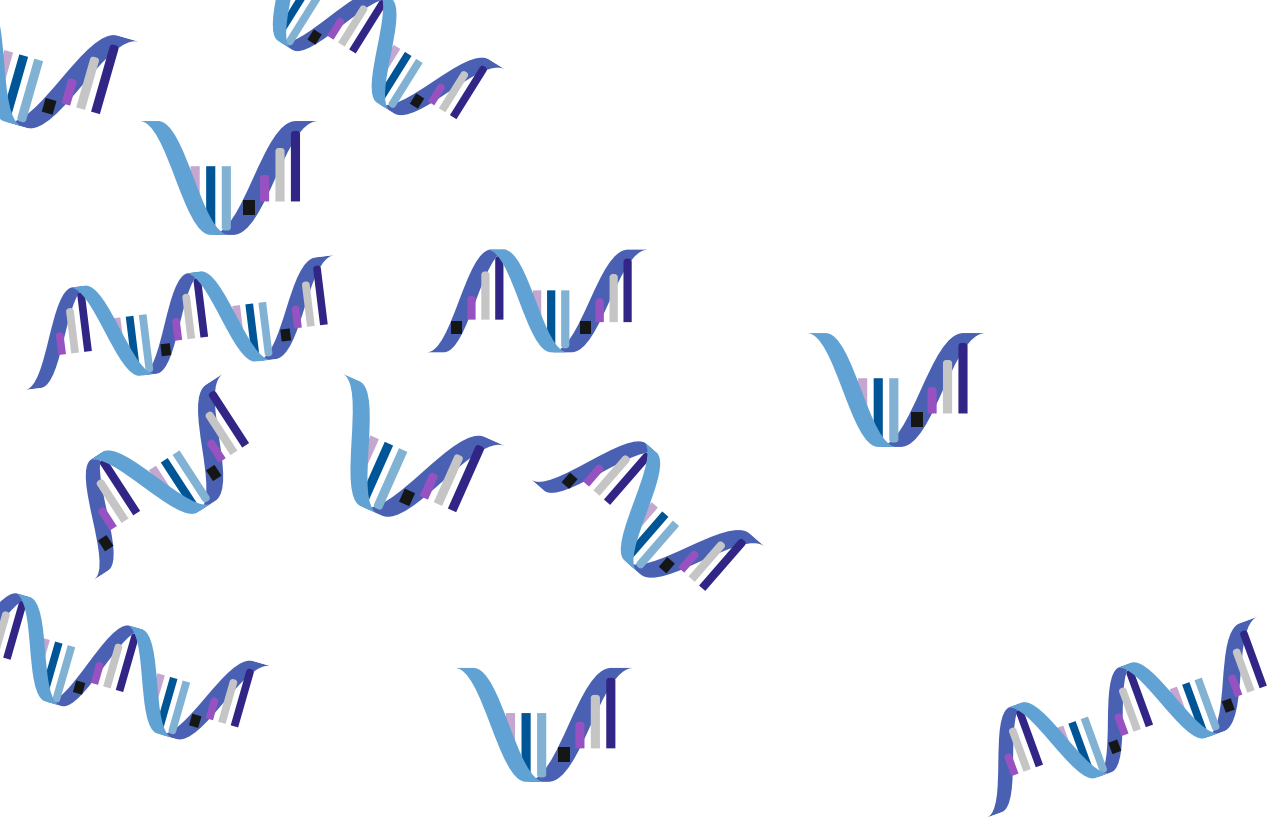
Multiple therapies are currently available for patients who are diagnosed with stage IV lung cancer. Since the first FDA approval of pembrolizumab in October 2015, more data have become available from the KEYNOTE studies on pembrolizumab (Figure 2), including data concerning the role of the PD-L1 biomarker. Most important are the data of the KEYNOTE-024 trial, published in October 2016, a first-line phase III trial which only included patients with a TPS ≥50%. With an ORR of 44.8% in the pembrolizumab group, compared to 27.8% in the chemotherapy group, there seems to be some evidence PD-L1 is a good marker for clinical benefit, which lead to the second FDA approval. However, although the expression of PD-L1 with a TPS ≥50% was associated with a significantly improved clinical benefit compared to TPS <50%, there was no difference between >1% (PD-L1 positive) and <1% (PD-L1 negative) in all dose regimes of pembrolizumab (Table 3). This implies that even in PD-L1-negative patients, responses can be as high as in the intermediate group (TPS 1–50%). A difference between TPS <50% and TPS ≥50% groups is seen in the KEYNOTE-010 trial, where only patients who were PD-L1 positive, were included. Here, ORR improved from 18% to 29–30%, respectively. Noticeable, although the KEYNOTE-006 trial [42] for melanoma showed no difference between 2 and 10 mg/kg pembrolizumab, a difference in OS is seen between these two groups. The hazard ratio for 2 mg/kg was 0.71, improving to 0.61 with the use of 10 mg/kg, suggesting10 mg/kg might be an option for patients who are <50% positive for PD-L1. Nevertheless, the predictive value of PD- L1 staining remains suboptimal and its use in clinical practice is questionable, considering the relatively high false-positive and false-negative rates. Therefore, the search for better predictive biomarkers for response to pembrolizumab and other immunotherapeutic drugs should continue. An interesting theory, recently published by Blank et al. [62], introduced the paradigm 'The Immunogram', in which simple blood tests (e.g. CRP and LDH) predict the response to immunotherapy. Another suggestion, not yet published, is that patients whose tumor show rapid growth may not be suitable candidates for immune-oncologic treatment. Both theories need further research. A phenomenon seen in immunotherapy is pseudo progression. Due to infiltration by activated T cells, an increased tumor volume can be seen on CT imaging. The patient might then be wrongly classified as progressive and treatment might be discontinued. This phenomenon complicates response evaluation by CT imaging, although recent data from our institute show less pseudo progression than expected, with a prevalence of only 2% [24, 157]. An early response marker would be helpful and more research is needed.

7. 5 year view

The first trial with pembrolizumab included patients in 2012 and at this moment most clinical trials of monotherapy with pembrolizumab are published [39, 40, 42, 133], which proved itself by leading to two approvals by the FDA for patients with advanced NSCLC. However, the first approved immune therapy for NSCLC was nivolumab, which is now widely used in many trials and seen as first-choice anti-PD-L1 therapy. The use of nivolumab does not require PD-L1 testing, which makes it the biggest competitor of pembrolizumab. However, the treatment schedule of pembrolizumab is less intensive

(Q2W for nivolumab versus Q3W for pembrolizumab), with comparable survival and quality of life (mostly compared to chemotherapy). Pembrolizumab is a welcome addition to the arsenal of anticancer drugs and has the potential to become first choice of treatment for many NSCLC patients in the near future. Results of clinical trials on other checkpoint inhibitors, such as durvalumab, atezolizumab, and avelumab are promising and should be awaited as well. Most promising for the nearest future are the results of the combination therapies. Data from the first trial, KEYNOTE-021 [135], were recently published in The Lancet. The combination of chemotherapy with pembrolizumab was compared to chemotherapy alone. The ORR was 55%, higher than 44.8% among patients with TPS ≥50% group in the KEYNOTE-024 trial, which is promising for other combination therapies. However, two other cohorts in the KEYNOTE-021 trial, evaluating the efficacy of the combination of pembrolizumab with ipilimumab, showed disappointing results, with an ORR of 24% (all 45 treated patients) and 17% (TPS ≥50%, 6 patients). In Table 2, other ongoing trials are shown. The most important are KEYNOTE-189 and KEYNOTE-407, both evaluating the combination of pembrolizumab with chemotherapy in a randomized controlled trial. Results are expected in March 2019 and hopefully in 5 years' time a successful combination is found. One important question for the nearest future concerning current immunotherapy is the duration of the therapy itself.

Currently, the golden standard is to treat patients for 2 years with pembrolizumab. No maximal treatment time is defined for nivolumab. Based on the theory of stimulating and 'learning' of the immune system, 12, 6, or even three months of treatment may be sufficient to reach and maintain a durable response. Shorter treatment times are found with other monoclonal antibodies. Standard treatment with ipilimumab for example, consists of only four cycles. Currently, the longest follow-up time is 3 years [149], so data on this subject is scarce. In 5-years' time, this question will be answered.



3 Blood platelet RNA profiles do not enable for nivolumab response prediction at baseline in patients with non-small-cell lung cancer

Mirte Muller*, Myron G. Best*, Vincent van der Noort, T. Jeroen N. Hiltermann, Anna-Larissa N. Niemeijer, Edward Post, Nik Sol, Sjors G.J.G. In 't Veld, Tineke Nogarede, Lisanne Visser, Robert Schouten, Daan van den Broek, Karlijn Hummelink, Kim Monkhorst, Adrianus J. de Langen, Ed Schuuring, Egbert Smit, Harry J.M. Groen, Thomas Wurdinger, Michel M. van den Heuvel

Tumor Biology 46 (2024) S327–S340

Abstract

Background: Anti-PD-(L)1 immunotherapy has emerged as a promising treatment approach for non-small-cell lung cancer (NSCLC), though the response rates remain low. Pre-treatment response prediction may improve patient allocation for immunotherapy. Blood platelets act as active immune-like cells, thereby constraining T-cell activity, propagating cancer metastasis, and adjusting their spliced mRNA content.
Objective: We investigated whether platelet RNA profiles before start of nivolumab anti-PD1 immunotherapy may predict treatment responses.
Methods: We performed RNA-sequencing of platelet RNA samples isolated from stage III-IV NSCLC patients before treatment with nivolumab. Treatment response was scored by the RECIST-criteria. Data were analyzed using a predefined thromboSeq analysis including a particle-swarm-enhanced support vector machine (PSO/SVM) classification algorithm.
Results: We collected and processed a 286-samples cohort, separated into a training/evaluation and validation series and subjected those to training of the PSO/SVM-classification algorithm. We observed only low classification accuracy in the 107-samples validation series (area under the curve (AUC) training series: 0.73 (95%-CI: 0.63-0.84, n=88 samples), AUC evaluation series: 0.64 (95%-CI: 0.51-0.76, n=91 samples), AUC validation series: 0.58 (95%-CI: 0.45-0.70, n=107 samples)), employing a five-RNA biomarker panel.
Conclusions: We concluded that platelet RNA may have minimally discriminative capacity for anti-PD1 nivolumab response prediction, with which the current methodology is insufficient for diagnostic application.

Keywords: Blood platelets, RNA-sequencing, nivolumab, immunotherapy

Introduction

Lung cancer, especially non-small cell lung cancer (NSCLC), is the most common cause of cancer-related mortality in the world, causing over one million deaths annually [66]. **I**nterestingly, PD-(L)1 checkpoint inhibitors have shown remarkable effectiveness, durability of response and tolerability [158], although overall response rates remain low (~20%) [3, 5]. **N**umerous studies have identified correlations between smoking, tumor tissue mutational load, infiltration of CD8-positive T-cells and response to anti-PD-(L)1 immunotherapy [159, 160]. **A**t the moment, current stratification of patients for anti-PD-(L)1 targeted therapy is hampered by limited accuracy and concordance of available biomarkers, including the currently used tumor tissue PD-L1-protein expression marker. Identification of patients with a low likelihood of response to anti-PD-(L)1 immunotherapy, while still correctly identifying individuals who most likely benefit from this therapy, may prevent unnecessary treatment and concomitant costs, and potential exposure of patients to serious immunological adverse events.

Blood platelets are cell fragments involved in hemostasis, initiation of wound healing and metastasis of cancer [161, 162]. They originate from the megakaryocytes that reside in the bone marrow and lung parenchyma [161, 163], and are loaded with pre-mature RNA molecules [164-166]. Upon environmental cues, platelets splice their mRNA molecules resulting in platelet RNA repertoires of both spliced and unspliced mRNAs [59, 164, 167]. In addition, platelets have the capability to sequester RNAs [168] and load platelet microparticles with RNAs that can be transferred to other cells, such as endothelial and tumor cells [169, 170]. Functionally, platelet-derived TGFß and direct physical platelet-tumor cell interactions can induce tumor cell epithelial-mesenchymal transition [171], a process that has previously also been associated with resistance to anti-PD1 immunotherapy [63]. In addition, platelets can behave as immunomodulators in inflammatory conditions [172, 173], regulate immune cell extravasation [174, 175], and cross-communicate with lymphocytes [176] and neutrophils [177]. It was shown that platelets armed with immunotherapeutic antibodies home to lung tumors in mice [178], exhibiting profound anti-tumor activity. Also, previously pretreatment increased platelet-to-lymphocyte ratio in patients with metastatic NSCLC treated with nivolumab was associated with worse overall survival and a reduced response rate [179]. Platelets are thus potentially involved in the immune response towards tumor cells, and might thereby alter their RNA composition.

We recently provided evidence that platelets may serve as blood-based liquid biopsies for the detection and subclassification of cancer [59, 61, 168], as well as for therapy response monitoring [180, 181]. In this prospective study, we investigated whether platelet RNA signatures may provide classification power for nivolumab immunotherapy response prediction before start of treatment.

Methods

Study design
In this prospective, observational, multi-center study in patients with advanced (stage III-IV) NSCLC blood samples at moment of baseline for nivolumab treatment were collected. The aim was to optimally identify patients who respond to nivolumab, for which clinical imaging data was available as a gold standard for response evaluation. Sample size calculation was performed using the method of Dobbin and Simon [182], assuming an overall response rate of 17% [2], employing the results from a pilot study (data not shown). We estimated a required samples size for the training series of 180 patients (i.e. ~31 responders and ~149 non-responders) under assumption that the gene with maximal mean fold difference between groups will exhibit a fold change of 2.5. From a clinical perspective we expected a successful biomarker to have 95% sensitivity and at least 50% specificity, i.e. halving the number of unnecessarily treated patients while keeping the number of false negatives low. We define the new biomarker to be successful if the lower boundary of the 95%-confidence interval for the specificity lies above 38% at the threshold/cut-off where its specificity is 95%. The rationale for this is that at the 95% sensitivity level a specificity of 38% will yield an overall response rate (ORR) in the positively tested group of 24% which is the rate achieved by the tissue PD-L1-test [2] at the much lower sensitivity of 46%. A validation series of 90 patients (75 non-responders and 15 responders) has 90% power to achieve this assuming a true specificity of the test of 55%.

Patient and sample collection
All patients with advanced (stage III/IV) NSCLC who were scheduled to start treatment with nivolumab were recruited for voluntary inclusion in this study. Patients were included in the Amsterdam UMC, Amsterdam, the Netherlands Cancer Institute (NKI/AvL), Amsterdam and the University Medical Center Groningen, Groningen, all from The Netherlands. Patients were included from December 2013 till June 2017, with follow-up until June 2018. The median time of follow-up was 195 days (Interquartile range (IQR) 97 – 399.5 days). All patients received treatment in accordance with recent literature [5, 183] and local guidelines. Platelet samples were collected up to one month before start of treatment, but not after start of treatment. Other clinical data, including computed tomography (CT) scans, blood tests and relevant clinical data such as the use of medication used in this study were collected for routine clinical practice. These were handled according to the Dutch Personal Data Protection Act (WPB). Response assessment of patients treated with nivolumab was performed by CT-imaging at baseline, 6-8 weeks, 3 months and 6 months after start of treatment. Treatment response was assessed according to the updated RECIST version 1.1 criteria and was scored as progressive disease (PD), stable disease (SD), partial response (PR), or complete response (CR) [105, 184]. A responder was defined as those patients with PE or CR at three monhts, despite possible PD at six months. A patient was also regarded a responder with SD at three and six months.

Tumor tissue was stained using either the PD-L1 22C3 (Agilent) or PD-L1 28-8 (Abcam) antibodies and scored according to the manufacturer's instructions. This study was

conducted in accordance with the principles of the Declaration of Helsinki, and all participant provided written informed consent. Our study was approved by the local medical ethical committee (PTC NKI-AvL, Amsterdam, NL45524.031.13; medical ethics committee, UMCG, Groningen, 2010/109 [185]; Dutch trial register NL7839) and the patient privacy committee, and was performed according to the institutional patient privacy protocols.

Platelet isolation

Peripheral whole blood was drawn by regular venipuncture in EDTA-coated Vacutainer tubes. Whole blood samples in 6- or 10-mL EDTA-coated Vacutainer tubes were processed using standardized protocols within 12 hours as described previously [168, 186, 187] at room temperature (21ºC). Here, using a 20-minute 120xg centrifugation step platelet rich plasma (PRP) was separated from nucleated blood cells. Following, 9/10th of the PRP needs to be collected carefully without disturbing the buffy coat, with the risk of contaminating the platelet preparation with nucleated cells. Then the platelets were pelleted from the PRP by a 20-minute 360xg centrifugation step. Platelet pellets were carefully resuspended in RNAlater (Life Technologies), followed by an overnight incubation at 4°C frozen at -80°C.

Platelet RNA-sequencing

The thromboSeq wet-lab platelet mRNA-sequencing protocol and dry-lab software modules have recently been described in detail and are publicly available [186]. In brief, first platelet total RNA was isolated using the mirVana miRNA isolation kit (Ambion, Thermo Scientific, AM1560). All samples were subjected to cDNA synthesis and amplification using the SMARTer Ultra Low RNA Kit for Illumina Sequencing v3 (Clontech, cat. nr. 634853) with an input of 500 picogram of total RNA. All amplified platelet cDNA was subjected to nucleic acid shearing by sonication (Covaris Inc) and subsequently labeled with single index barcodes for Illumina sequencing using the Truseq Nano DNA Sample Prep Kit (Illumina, cat nr. FC-121-4001). All bead clean-up steps were performed using a 15-minute bead-cDNA binding step and a 10-cycle enrichment PCR. Other steps were according to manufacturer's protocol. During this protocol, multiple quality control steps were performed using RNA and DNA chips available for the Bioanalyzer 2100 (Agilent). High-quality samples with product sizes between 300-500 bp were pooled (12-24 samples per pool) in equimolar concentrations for superficial thromboSeq and submitted for 100 bp Single Read sequencing on the Illumina Hiseq 2500 and 4000 platform. During whole blood, blood platelet and platelet RNA processing, the researchers were blinded to the clinical outcome measures.

Processing of raw RNA-sequencing data

We employed Trimmomatic (v. 0.22) [188] for trimming and clipping of sequence adapters from the RNA-seq reads, of which the resulting reads were mapped to the human reference genome (hg19) using STAR (v. 2.3.0) [189]. The mapped reads were summarized using HTSeq (v. 0.6.1), guided by the Ensembl gene annotation version 75 [190]. All following statistical and analytical analyses were performed in R (version 3.3.0) and R-studio (version 0.99.902). Genes encoded on the mitochondrial DNA and the Y-chromosome were excluded from downstream analyses. To circumvent potential

contamination from cell-free DNA, only intron-spanning spliced RNA reads were employed for data analysis [60]. Sample filtering was performed by assessing the library complexity, which is partially associated with the intron-spanning reads library size. The raw sequencing data has been deposited at the GEO database under accession number GSE216297.

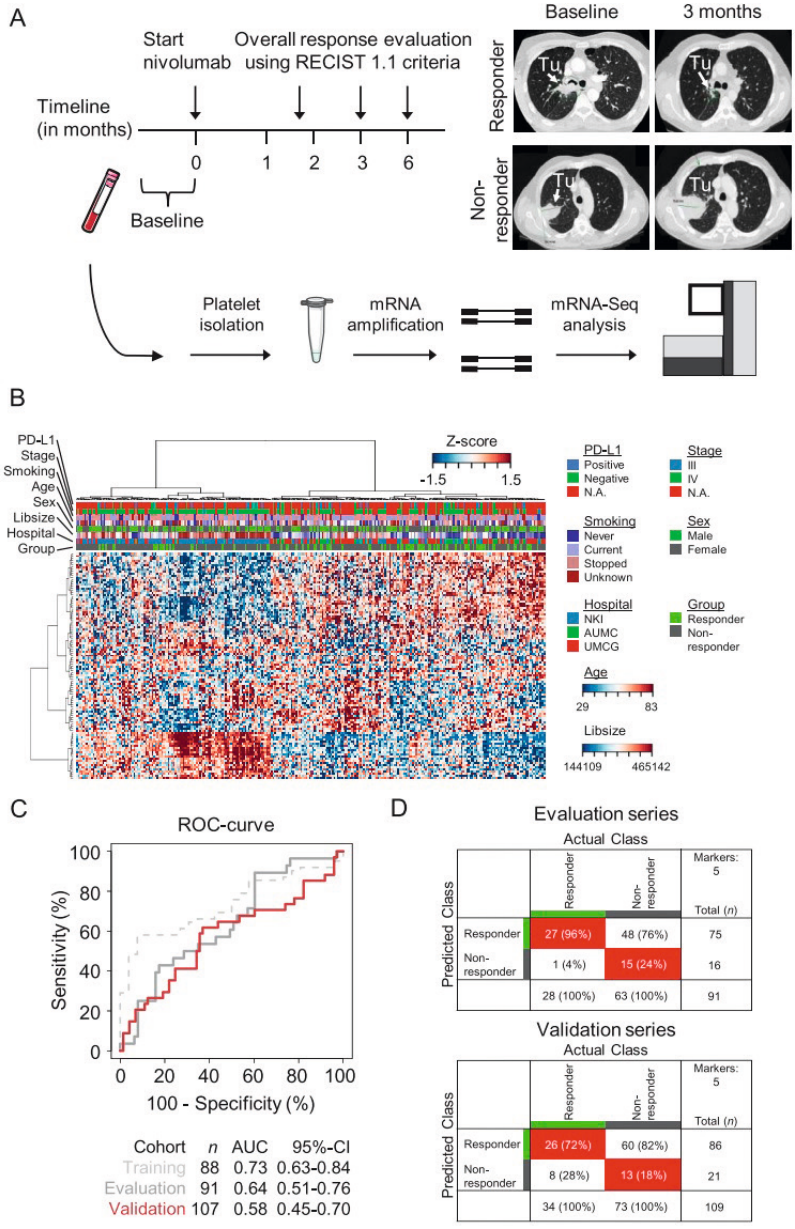


Figure 1 – Platelet RNA-based nivolumab response prediction.

ThromboSeq classification algorithm

We employed our previously standardized and optimized thromboSeq algorithm for data analysis [59-61], which is extensively described elsewhere. For this study, software-version v1.5 of the thromboSeq source code was employed, and is for research purposes publicly available via GitHub (https://github.com/MyronBest/thromboSeq_source_code_v1.5) [61]. In short, the particle-swarm-optimization (PSO-) enhanced support vector machine (SVM)-based thromboSeq algorithm employs an SVM-algorithm for RNA expression classification tasks, supplemented by a PSO algorithm that iteratively selects parameters for gene panel selection. The thromboSeq algorithm is developed with the use of a training, evaluation, and independent validation series. The training series is used for selection of stable genes that are used for remove unwanted variation (RUV)-mediated data normalization, gene panel selection and SVM-training. Here, RUV is employed to minimize variability among the samples introduced by potential confounding factors [60]. This correction measure is determined based on the training series to enable for independent validation of remaining samples. The gene panel is selected using ANOVA-statistics, supplemented by filtering for highly correlated RNAs as well as filtering for RNAs most contributive to the SVM-model. Based on this selected gene panel a SVM-model including class weight correction is trained. Next, the evaluation series is employed to evaluate the classification performance of the SVM-model with the particular gene panel and algorithm settings. Here, a PSO-algorithm is used to repeatedly provide more optimized thresholds to select the gene panel and set other algorithm settings, of which the optimization steps are evaluated using the classification scores in the evaluation series. Finally the algorithm and gene panel is locked following which the validation series is used for validation of the PSO-optimized SVM-classification algorithm.

Figure 1 – Platelet RNA-based nivolumab response prediction. *(continued)*

(A) Schematic overview of the experimental setup. Blood of patients eligible for treatment with PD-1 inhibitor nivolumab was included from one month before till start of treatment (baseline, t=0). Tumor response read-out based on CT-imaging and according to the RECIST 1.1 criteria were performed at 6-8 weeks, 3 months, and 6 months, after start of nivolumab therapy. A responder was defined as PR, CR at three months and SD at three and six months. A non-responder was defined as PD and SD after three, but PD after six months.

(B) Heatmap and unsupervised hierarchical clustering analysis of RNAs with differential RNA levels between nivolumab responders (n=88) and non-responders (n=198). Columns indicate samples, rows indicate RNAs, and color intensity represents the Z-score transformed RNA expression values. Clustering of samples showed non-random partitioning ($p<0.003$, Fisher's exact test). The clinical and technical variables 'hospital of sample location', 'library size', 'sex', 'age', 'smoking status', 'tumor stage', and 'PD-L1 tumor tissue status' are indicated on top of the heatmap with each owns color coding (legend on the right).

(C) Receiver operating characteristics curve of the thromboSeq nivolumab response prediction algorithm at moment of treatment baseline. Indicated are the training series (dashed gray line), evaluation series (gray line), and validation series (red line). Also indicated are sample numbers, AUC values, and the 95% confidence intervals (95%-CI).

(D) 2x2 tables summarizing the classification results for the evaluation series (upper table) and validation series (lower table) of the thromboSeq nivolumab response prediction algorithm at moment of treatment baseline geared towards a rule-out setting. Indicated are sample numbers and percentage of total samples.

For this study, the variables 'Age' and 'lib.size' were applied as potential confounding factors for RUV-correction with both a default threshold of 0.8. The swarm-parameters 'lib.size', 'FDR', 'correlatedTranscripts', and 'rankedTranscripts' were employed using the following boundaries, respectively: -0.1 – 1.0; 5 – number of genes with ANOVA FDR <0.5; 0.5 – 1.0; 5 – number of genes with ANOVA FDR <0.5. In total 100 particles were deployed with eight iterations. Class-weights and a rule-out optimization step forcing towards at least 95% sensitivity in the evaluation series during particle optimization
were enabled. The algorithm employed has selected the following thresholds: 'lib.size': 0.131965039293101; 'FDR': 5; 'correlatedTranscripts': 0.989771223666639; 'rankedTranscripts': 12.2922697840516.

An ANOVA comparison was performed using the thromboSeq.anova-function in the thromboSeq software. The variables 'Age' and 'lib.size' were subjected to RUV-correction, with the thresholds '0.8' and '0.8', and with k.variables=3. The heatmap clustering was optimized by a PSO-function thereby adjusting the FDR-threshold evaluated by column-dendrogram clustering.

Statistical analyses

SPSS (v25; SPSS, Chicago, USA) was used for the descriptive statistics. For a between-group-comparison the Mann-Whitney U test, two sample unpaired Students t-test, or chi-square test was used depending on the variable.

Results

We prospectively collected 300 blood samples from patients with advanced NSCLC that were selected for treatment with the PD-1 monoclonal antibody nivolumab. These were subjected to the thromboSeq protocol (Figure 1A). 14 samples (4%) were excluded from follow-up analyses according to previously introduced thromboSeq quality measures (Supplemental Figure 1) [60, 186]. The remaining 286 samples were included for the analyses (Table 1). We observed no statistically significant differences between responders and non-responders regarding age, gender, and hospital of patient inclusion (Table 1). We also observed no statistical significant differences in total platelet counts between responders and non-responders of whom this data was available (p=0.660, n=185, unpaired Student's t-test, Table 1). Tumor tissue was partially available for correlative PD-L1 immunohistochemistry analysis. Additional demographic variables are presented in Table 1 and Table S1. The full dataset was subdivided into training (n=88), evaluation (n=91), and validation series (n=107; Supplemental Figure 1C), as previously performed according to our thromboSeq dry-lab protocol [186]. For comparative analysis, we forced the patients whom the tumor tissue PD-L1 status was available into the validation series. The hospital of origin (i.e. sample collection) was equally distributed among the three series (Table S2). The separation of samples among the three series was defined in the study protocol and based on preliminary data and power calculations (see Methods).

We first performed an ANOVA comparison on all platelet RNA samples, identifying platelet RNAs with discriminative power between nivolumab responders and non-responders. From the 3750 platelet RNAs that were detected in the dataset in total,

27 had statistically significant differential RNA levels (False Discovery Rate (FDR)<0.05, Table S3). From these RNAs, only one RNA from the GSEA inflammatory response gene data set [191] had a statistically significant differential expression level level (NAMPT, FDR = 0.0002).

Heatmap dendrogram clustering of the samples applying these 27 RNAs resulted in non-significant clustering between responders and non-responders (p-value: 0.24, Fisher's exact test, data not shown). Adjusting the ANOVA-statistics significance threshold employing a PSO-function (see Methods) resulted in dendrogram clustering discriminating responders from non-responders (p-value<0.003, Fisher's exact test, used FDR-threshold: 0.419696, Figure 1B). Improving discriminating power by adjusting the ANOVAs FDR-threshold implies that also statistically non-significant RNAs can contribute to the clustering accuracy. This provides the rationale for the PSO-enhanced classification algorithm development selecting an RNA panel based on most contributive power to a discriminative biomarker panel.

We next performed PSO-enhanced training of an SVM-classification algorithm, forcing the algorithm to identify optimal algorithm settings including the RNA biomarker panel for a rule-out nivolumab response prediction classifier. This would enable accurate identification of patients who do respond to nivolumab, while providing the opportunity to exclude patients that will not respond. This training process resulted into an area under the curve (AUC) for the training series of 0.73 (95%-CI: 0.63-0.84, n=88 samples, as determined by LOOCV), and for the evaluation series of 0.64 (95%-CI: 0.51-0.76, n=91 samples) for the evaluation series. Subsequent validation of the 107-samples validation series resulted into an AUC of 0.58 (95%-CI: 0.45-0.70, n=5 spliced RNA biomarker panel, Figure 1C). The five-gene biomarker panel as unbiasedly selected by the thromboSeq software was composed of the RNAs HPSE, HBD, PRSS50, CRYM, and LARP1 (Supplemental Figure 2A). Using the classification score readout according a 'rule-out' application, in a sensitivity of 96% (95%-CI: 82-100%, n=28) with a specificity of 24% (95%-CI: 14-36%, n=63) in the evaluation series, resulted in a concomitant sensitivity in the validation series of 76% (95%-CI: 59-89%, n=34) and a specificity of 18% (95%-CI: 10-29%, n=73, Figure 1D), with which our predefined threshold of a successful new biomarker is not reached.

Discussion

Anti-PD-(L)1 therapies have revolutionized cancer treatment, although, (non-invasive) predictive biomarkers are highly desirable [192]. In this proof-of-concept study, we aimed to associate spliced platelet RNA profiles to nivolumab response prediction measures at moment of therapy baseline.

Platelets act as local systemic responders during tumorigenesis [162], thereby suffering from tumor-mediated platelet education, and altering its behavior [171, 174, 193]. In previous studies, we have demonstrated that platelet RNA can function as a biomarker trove to detect and classify cancer from blood [59, 61]. Using our previously developed thromboSeq platform, we could not find solid RNA repertoires associated with tumor

response at moment of start of therapy. Therefore, we were in this setting and study setup unable to successfully validate a nivolumab response prediction thromboSeq algorithm. Therefore at this point, interrogation of platelet RNA signatures before start of nivolumab treatment seems not to be useful for nivolumab therapy selection.

Table 1 - Baseline characteristics of responders versus non-responders

	Non-responder N=198	Responder N=88	Total (N=286)	P-value
Patient				
Male sex – No. (%)	115 (58.1)	43 (48.9)	158 (55.2)	0.15 [A]
Age (yr) – Mean (95% CI)	63.8 (62.5-65.1)	64.9 (62.8-67.1)	64.1 (63.0-65.3)	0.61 [B]
Smoking status – No. (%)				**0.05** [C]
Never	20 (10.1)	4 (4.5)	24 (8.4)	
Current	33 (16.7)	25 (28.4)	58 (20.3)	
Stop	132 (66.7)	57 (64.8)	189 (66.1)	
Pack Years – Mean (95% CI)	32.6 (29.6-35.6)	37.4 (33.8-41.1)	34.2 (31.9 – 36.5)	**0.01** [B]
WHO Performance status ≥ 2 – no / total known patients	31/182	2/84	33/266	**0.003** [C]
Platelet not normal range [A*] – no. / total known patients	27/128	7/57	34/185	0.15 [A]
Platelet Count Baseline – Mean (95% CI)	338.0 (314.5-361.6)	326.7 (295.4-357.9)	334.5(315.7-353.3)	0.66 [B]
Brain metastasis	43	17	60	0.65 [A]
Pathology – no.				
Adenocarcinoma	141	60	201	0.46 [A]
Squamous cell carcinoma	47	22	69	
Other	7	6	13	
Unknown	3	0	3	
PD-L1 – no.				
Negative	26	12	38	0.48 [C]
Positive 1 – 50%	10	5	15	
Positive >50%	3	4	7	
Unknown	159	67	226	
Previous treatment – no.				
CCRT	19	10	29	0.86 [A]
Platinum chemotherapy	109	47	156	
Unknown	67	29	96	
Line of treatment – no.				
1	2	1	3	0.58 [A]
2	101	43	144	
<2	25	14	39	
Unknown	70	30	100	
Previous RT / total known patients	87/127	34/58	121/185	0.24 [A]
Thoracic RT / total known patients	53/127	26/58	79/185	0.75 [A]

Table 1 - Baseline characteristics of responders versus non-responders *(Continued)*

	Non-responder N=198	Responder N=88	Total (N=286)	P-value
Platelet related medication – no.				
None	82	38	120	0.43 [A]
Acetylsalicyclic acid	23	9	32	
Other	22	11	33	
Unknown	71	30	101	
Hospital – no.				
NKI	90	38	128	0.84 [A]
AUMC	39	20	59	
UMCG	69	30	99	
Survival				
OS – mean (95% CI)	165.3 (146.1-184.55)	505.8 (465.7-545.8)	270.4(244.7-296.2)	<0.001 [B]
PFS – mean (95% CI)	72.6 (59.4-85.8)	455.8 (413.3-498.4)	190.5 (164.5-216.5)	<0.001 [B]

Abbreviations: N: Number of patients; no.: Number of patients; yr: years; 95% CI: 95% confidence interval; WHO performance-status score: World Health Organization performance status score. this is a score ranging from 0 to 5. where 0 indicates no symptom. 1 indicates mild symptoms and above 1 indicates greater disability; PD-L1: Programmed death ligand 1; CCRT: chemoradiation therapy; RT: radiation therapy; RECIST: Response evaluation criteria in Solid Tumors; NKI: Netherlands Cancer Institute. AUMC: Amsterdam University Medical Centers. UMCG: University Medical Center Groningen. OS: Overall Survival; PFS: progression free survival.* Either below under limit of normal or above limit of normal. as defined by the clinical laboratory. In the Netherlands a normal platelet count is between 150-350 platelets x 10^9 / liter (150.000-350.000 platelets per microliter). Used statistical tests: [A]: Chi Square, [B]: One way ANOVA, [C]: Fisher-Freeman Halton exact test.

Subsequent analysis of classification accuracies according to PD-L1 tumor tissue status of whom the data was available showed on average increased classification scores for patients responding to nivolumab as compared to patients not responding to nivolumab (Supplemental Figure 2B).

Because platelets are regarded as local immuno-like responders, the role of platelets in anti-tumor activity prior to and during anti-PD-(L)1 therapy, including direct and indirect physical interactions with lymphocytes, macrophages, and tumor cells should not be underestimated. In our five-gene platelet biomarker panel, both HPSE (heparanase) and LARP1 (La RibonucleoProtein 1) have lower expression values in nivolumab responders. HPSE is active in tumor invasion and inflammatory processes [194, 195], whese-as LARP1 is associated with dengue virus and coronavirus disease-19 (COVID-19) [196, 197]. Higher expression levels in nivolumab non-responders may suggest either a different inflammatory process required for a anti tumor response, or a highly progressive tumor, which may explain the insensitivity to nivolumab treatment.

In this light, the current results were unexpected. We have several explanations for our findings; 1) platelet RNA is only very subtly altered in the presence of immunotherapy-sensitive as opposed to immunotherapy-resistant NSCLC tumors for which insufficient sample numbers were available to uncover such differences Aternatively, the employed

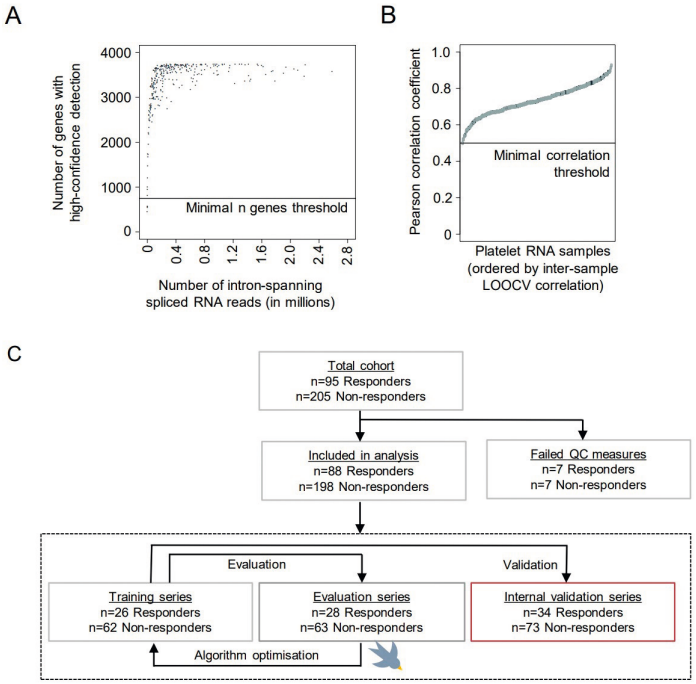
machine learning algorithms and gene panel selection approach had insufficient discriminatory power; 2) the alterations are masked by the presence of multiple platelet subpopulations; 3) the alterations may be best measured after start of treatment, e.g. a couple of weeks following the first nivolumab exposure; 4) insufficient samples were included into the training process; 5) despite matching of several confounding factors, remaining (unmeasured) factors and/or pre-analytical variables may increase the gene expression noise as opposed to the (low-level) signal and 6) in the power calculation the estimated RNA expression difference was overestimated, resulting into a too small anticipated cohorts sample size and therefore to a reduced algorithm's performance. Also, we cannot exclude that platelet RNA is of value as a predictive biomarker platform for other (immuno) therapeutics. Here, re-analysis of the data using other machine learning approaches is of interest and the data is publicly available for such purposes.

Acknowledgements
We are thankful to the NKI-AVL Core Facility Molecular Pathology and Biobanking (CFMPB) for supplying NKI-AVL Biobank material and lab support and Sebastiaan van de Sand (SIT B.V.) for computational resources. We also thank Maarten Slagter for his support. Financial support was provided by European Research Council E8626 (R.J.A.N., E.F.S., T.W.) and 336540 (T.W.), the Dutch Organisation of Scientific Research 93612003 and 91711366 (T.W.), the Dutch Cancer Society (T.W.), BMS IION (M.M.v.d.H., T.W.), and Stichting STOPHersentumoren.nl.

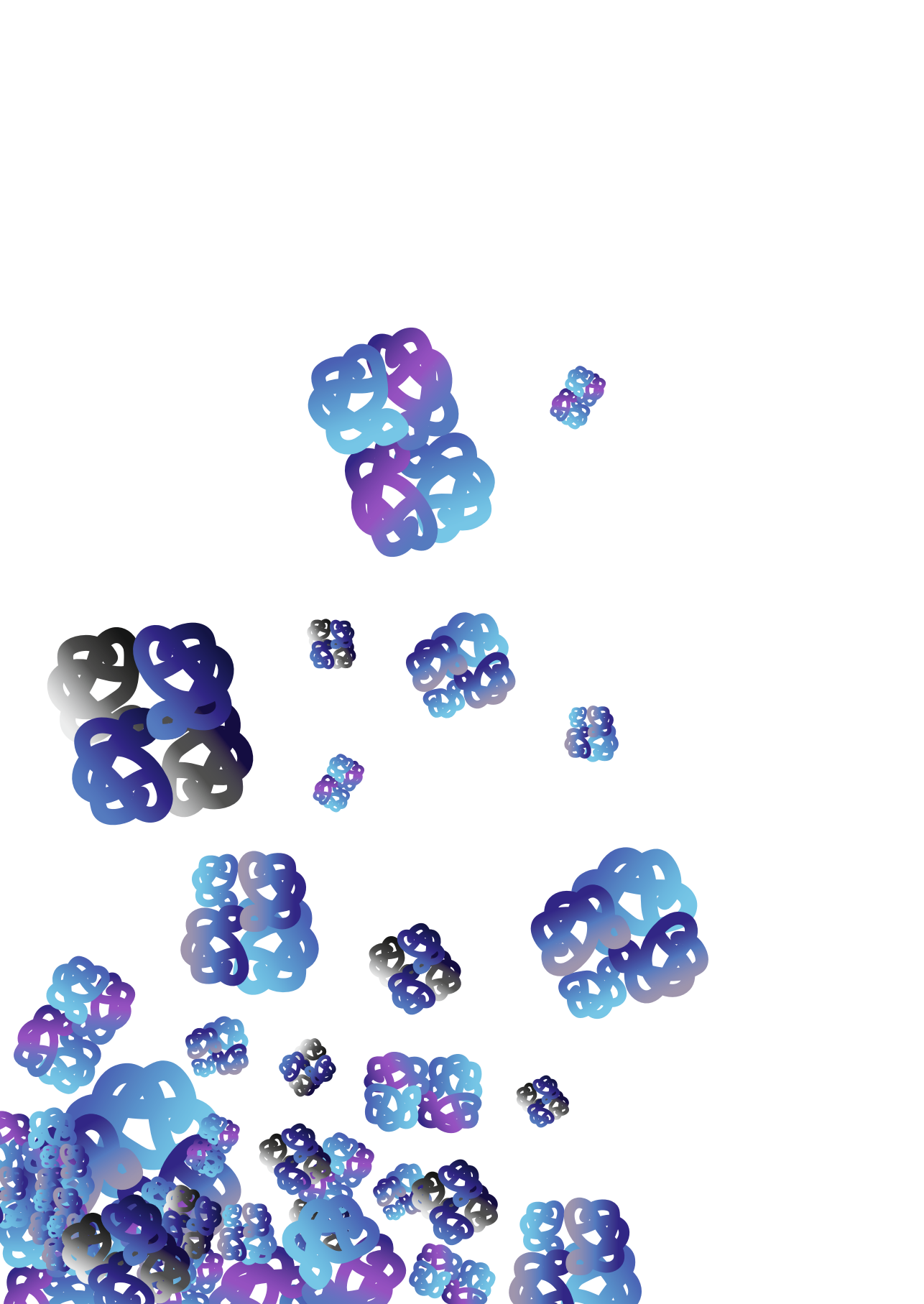
Supplemental material

The supplemental material can be found at:

Included content in this thesis:
Figure S1 - ThromboSeq bioinformatics QCs



Supplemental Figure 1 – thromboSeq bioinformatics QCs
(A) Number of intron-spanning spliced RNA reads versus number of genes detected. Samples with at least 1,500 different RNAs detected were included for further analyses.
(B) Leave-one-sample out cross-correlations, assessing the intersample comparability (Pearson correlation) based on all detected RNAs and filtering out samples with low inter-sample correlation (minimal correlation threshold >0.5).
(C) Schematic flow diagram indicating the number of samples at start of the analysis, following *in silico* QC-filtering, and available for algorithm training, evaluation and validation.



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

M. Muller*, K. Hummelink*, D.P. Hurkmans, A.N Niemeijer, K. Monkhorst, J. Roder, C. Oliveira, H. Roder, J.G. Aerts, E.F. Smit

Clin Cancer Res October 1 2020 (26) (19) 5188-5197

Abstract

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

Experimental Design: In a retrospective study, mass spectrometry (MS)-based proteomic analysis was performed on pretreatment sera derived from patients with advanced NSCLC 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 was 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 two 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% confidence interval, 0.38–0–87; *P* = 0.009). There was no significant association with clinical factors including PD-L1 expression, available from 133 of 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 patients with advanced NSCLC treated with second-line PD-1 CPIs and deserves further prospective study.

Translational Relevance

Predictive biomarkers for the efficacy of PD-L1 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 and overall survival (OS) across two validation cohorts (*N* = 98 and *N* = 75). In pooled analysis, a significantly better OS was demonstrated for "sensitive" relative to "not sensitive" patients, HR 0.58 (95% confidence interval, 0.38–0.87; *P* = 0.009). There was no significant association with clinical factors including PD-L1 IHC. Further prospective exploration of the predictive capabilities of this assay is underway.

1. 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% [198]. In recent phase III studies, either comparing CPIs alone or CPI chemotherapy to chemotherapy[199], 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 [40, 200].

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-L1 expression [40], staining of CD8-positive cells[201], tumor mutational burden (TMB) [202], and other genomic markers [203, 204]. The predictive power of the best studied of these PD-L1 IHC is far from perfect. For example, in patients with previously treated NSCLC with PD-L1 staining of at least 50%, the objective response rate to pembrolizumab was 44% [40]. 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.

2. Materials and Methods

2.1 Patient cohorts and sample sets

Pretreatment serum samples, collected within 1 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 (Amsterdam, the Netherlands) between May 2015 and March 2017. Validation set 1 consisted of 98 samples from patients treated at the Vrije Universiteit Medical Center (Amsterdam, the Netherlands) or the Netherlands Cancer Institute (Amsterdam, the Netherlands) between June 2015 and July 2018. Validation set 2 comprised samples from 75 patients treated at the Erasmus University Medical Centre (Rotterdam, the Netherlands) between April 2016 and July 2017. Patients, identified according to criteria established in the phase III trials demonstrating benefit for nivolumab over docetaxel [3, 5], received nivolumab 3 mg/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 3 months thereafter. In addition, a fourth cohort of pretreatment serum samples (chemotherapy cohort) was collected from patients with advanced NSCLC treated in second-line with chemotherapy while enrolled in a clinical trial (NCT00989690)[205].

Samples were available for 68 of the 74 patients who received docetaxel (75 mg/m2 every 21 days) in this study. Trial inclusion and exclusion criteria have been published elsewhere [205]. 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.

2.2 PD-L1 IHC

Tumor PD-L1 expression scoring was performed according to the instruction manual of the qualitative IHC assay developed as a complementary diagnostic tool for nivolumab (PD-L1 IHC 22C3 pharmDx, Dako). 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 with PD-L1–stained slides 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.

2.3 Spectral acquisition and processing

Samples were processed using standardized operating procedures. We used the Deep matrix-assisted laser desorption/ionization (MALDI) method of mass spectrometry on a MALDI Time-of-Flight Mass Spectrometer (SimulTof Systems) to generate reproducible mass spectra from small amounts of serum (3 µL) [206]. 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 Supplementary Data). Sample processing and MS analysis followed methods presented previously [207, 208] and are outlined in the Supplementary Materials and Methods. 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.

2.4 Test development

Test development was carried out using the Diagnostic Cortex platform[209], which has been used previously to design tests that were able to stratify patients by outcome in various settings, for example, to identify patients with advanced melanoma likely to be sensitive to CPIs [207, 208]. The approach incorporates machine learning concepts and elements of deep learning[210] to facilitate test development in cases where there are more measured attributes than samples. The potential for overfitting was minimized, thus allowing the creation of tests that can generalize to unseen datasets. Tests were 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") [211].

For successful supervised learning, suitable training class labels are required. We used a semisupervised approach [212] that does not require accurate prespecification 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 was 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 was 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 were 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 [212]. Full details of the application of the method in this setting are provided in the Supplementary Materials and Methods.

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 [208]. It has been validated in multiple independent cohorts of patients with melanoma treated with CPIs [208, 213]. Also, it has demonstrated some ability to stratify OS of patients with advanced NSCLC treated with nivolumab [214]. A version of BDX008, adapted for the spectral preprocessing parameters and feature definitions in this project, was created (see Supplementary Data: Supplementary Materials and Methods for details).

Preliminary statistical considerations showed a binary split of the development set into equal-sized groups would allow detection of a HR 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 an HR 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.

2.5 Protein set enrichment analysis

This analysis applies the gene set enrichment analysis (GSEA) method [215] 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). Specific protein sets were created as the intersection of the list of the panel targets and results of queries for biological functions from gene ontology, using AmiGO2 tools (http://amigo.geneontology.org/amigo) and UniProt databases (https://www.uniprot.org/). The protein set enrichment analyses (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 [216]. 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 [216] 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[217]. The PSEA was carried out using a C# implementation and MATLAB (MathWorks). PSEA P values were defined as described by Subramanian and colleagues[213]. FDRs for the PSEA calculations were assessed using the method of Benjamini and Hochberg [218].

2.6 Other statistical analysis

All analyses, except the PSEA, were carried out using SAS9.3 (SAS Institute) or PRISM (GraphPad). Survival/PFS plots and medians were generated using the Kaplan–Meier method. Association between test classification and categorical or continuous variables was assessed using Fisher exact test and Mann–Whitney test, respectively. All P values are two-sided.

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 (%)	Adenocarcinoma	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 expression, 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)

Table 1 - Patient characteristics and outcomes for all cohorts: development, validation set 1, validation set 2, and chemotherapy cohort. *(Continued)*

		Development (N=116)	Validation 1 (N=98)	Validation 2 (N=75)	Chemotherapy (N=68)
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

3. 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. Clinicopathologic 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.

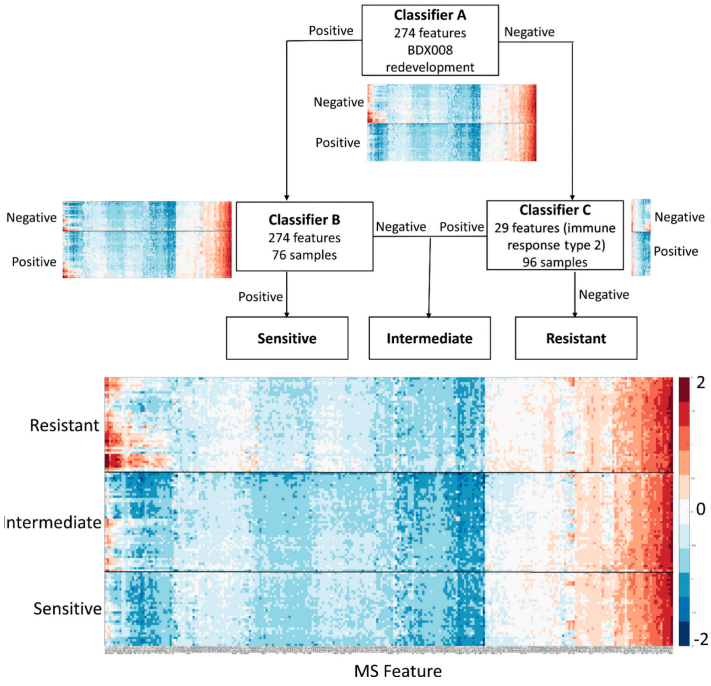


Figure 1: Schema showing how the final test result is produced from the three classifiers A, B, and C.

Notes: *Heatmaps within the schema 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).

3.1 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, that is, 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 preexisting 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 and Methods). 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 Supplementary Data. 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%.

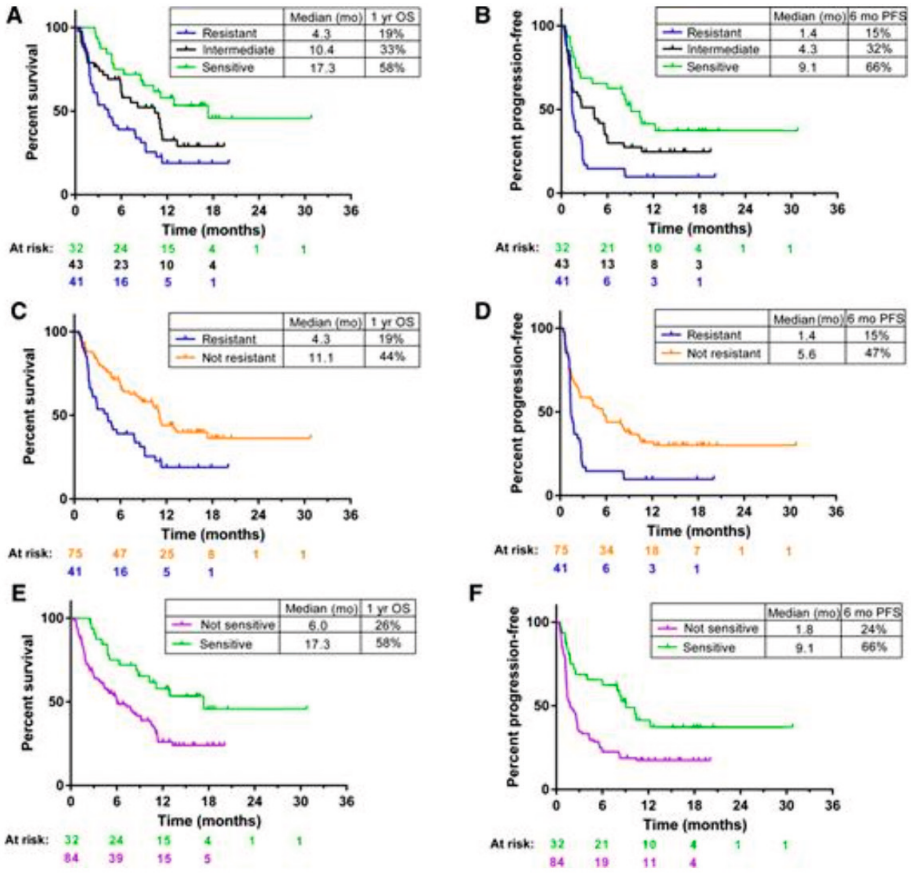


Figure 2 - Outcomes by test classification for the development cohort.

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 Figure 2A and 2B. PFS for the resistant subgroup was significantly shorter than for the other groups [resistant vs. sensitive: HR, 0.33 (95% confidence interval (CI), 0.19–0.58); P < 0.001 and 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 and 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 Supplementary Data: Supplementary Results; Supplementary Table S12). Eighty-five percent of patients classified as resistant experienced progressive disease as best response and only 10% had a response (all partial). In the sensitive group, only 28% of patients had progressive disease as best response and 28% achieved a response (one complete response (CR) and eight partial responses (PR) as best response 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, that is, the combination of intermediate and sensitive subgroups, see Figure 2C and D. The resistant subgroup had significantly inferior OS and PFS than the other subgroups [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, that is, the remainder of the cohort (resistant + intermediate subgroups; Figure 2E and F). 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 of sensitive versus not sensitive did not retain its independent significance as a predictive factor (Supplementary Data: Supplementary Results; Supplementary Tables S13 and S14).

Baseline patient characteristics showed no association with test classification for P < 0.05 (Supplementary Data: Supplementary Results; Supplementary Table S15). In particular,

PD-L1 expression was not significantly correlated with test classification (P = 0.387 for ternary classification vs. PD-L1+/PD-L1−/NA).

3.2 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 [219] 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 Supplementary Data.

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 versus not resistant and sensitive versus not sensitive, are shown for the validation sets in Figure 3A–3D. In validation set 1, Figure 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 1 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 Supplementary Data: Supplementary Results; Supplementary Figs. S1 and S2). 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 Figure 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 an 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 Supplementary Data: Supplementary Results. Analysis of PFS showed only numerical differences between classification groups.

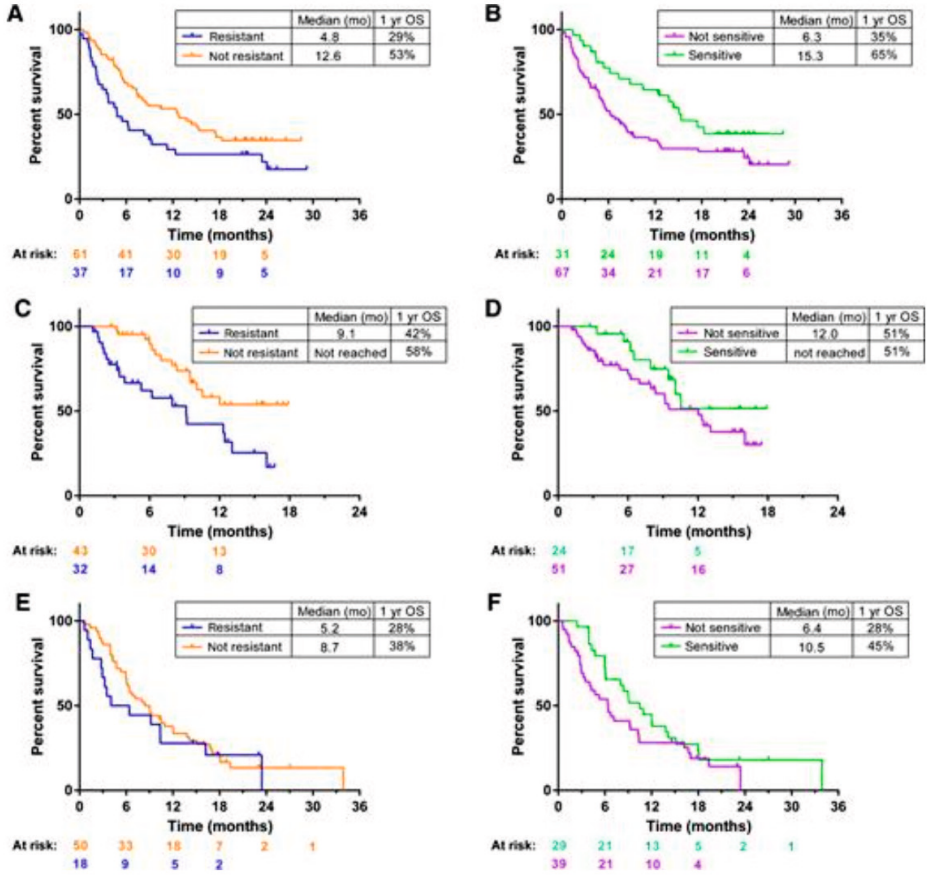


Figure 3 - Kaplan–Meier plots of OS for the validation sets and the chemotherapy cohort.
Notes: *A, Validation set 1 resistant versus not resistant. B, Validation set 1 sensitive versus not sensitive. C, Validation set 2 resistant versus not resistant. D, Validation set 2 sensitive versus not sensitive. E, Chemotherapy cohort sensitive versus not sensitive. F, Chemotherapy cohort resistant versus not resistant.*

As results were consistent across cohorts, within the limits of relatively small subgroup sizes, a pooled analysis of all patients treated in second-line setting 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, and 0.337 for ternary, resistant vs. not resistant, and 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 (95% CI, 1.01–2.54); P = 0.046 for OS and HR, 1.61 (95% CI, 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 versus 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 versus not sensitive

stratification was a significant independent predictor of OS (P = 0.009) and showed a trend for prediction of PFS (P = 0.079).

Figure 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).

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 value**	**HR (95% CI)**	**P value**
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 value**	**HR (95% CI)**	**P value**
Test (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

Table 3 - PSEA of test classifications resistant vs 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
Natural Killer (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
Type17 immune response	0.4668	<0.6
Type1 immune response	0.5102	<0.7
Type2 immune response	0.7791	<0.9
Response to hypoxia	0.9287	<1
T cell mediated immunity	0.9861	<1
IFN-Gamma	0.9884	<1

3.3 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[215-217]. 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 $P < 0.001$. In addition, 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 versus not sensitive phenotype are contained in the Supplementary Data: Supplementary Results; Supplementary Table S21.

4. 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. On the basis of 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 versus 11.1 months, respectively. The test was validated while blinded to clinical outcome data with an independent set of patients with advanced NSCLC, treated at a different institution. The classifier failed to stratify outcomes within a historical cohort of patients with advanced NSCLC 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 intrapatient tumor heterogeneity and could assess host factors that are not captured by examination of the tumor microenvironment in histologic 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 [220, 221]. Complement activation, as measured by Cd4, a stable complement degradation product, in serum of patients with early-stage NSCLC was significantly associated with poor prognosis [222]. 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[64]. 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, among them NSCLC [63]. Interestingly, sera derived from patients with tumors exhibiting wound-healing signatures elicited identical signatures from nontumor-associated fibroblasts, which were found to be a powerful predictor of an unfavorable clinical course [223]. 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.

The results obtained in this study do not stand alone. Weber and colleagues identified a protein classifier from sera of patients with melanoma 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 [207]. 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 patients with advanced NSCLC [224]. 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 versus docetaxel for OS and PFS ($P_{interaction} < 0.01$). 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 was low and all three immunotherapy-treated cohorts came 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 was not mandatory for administration of nivolumab in the second- and higher-line setting. Unfortunately, TMB data were 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 was 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 first-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.

Supplemental material

The supplemental material can be found at:

Incluced content in this thesis:
Supplemental results:
Supplementary Table 9: Classification concordance (sensitive vs intermediate vs resistant)
Supplementary Table 10: Classification concordance (resistant vs not resistant)
Supplementary Table 11: Classification concordance (sensitive vs not sensitive)
Supplementary Table 12: Response by test classification in the development cohort
Supplementary Table 13: Multivariate Analysis of OS and PFS resistant vs not resistant
Supplementary Table 14: Multivariate Analysis of OS and PFS sensitive vs not sensitive
Supplementary Table 19: PD-L1 status by test classification (pooled second line patients)
Supplementary Table 21: PSEA for association of biological processes
Supplementary Figure 5: Kaplan-Meier plots of OS by test classification
Supplementary Figure 8: Dot plot of PD-L1 staining by test classification
Supplementary Figure 9: Kaplan-Meier plots of OS by test classification

Supplemental Data: Results

5. 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 9 for classification sensitive vs intermediate vs resistant, in Supplementary Table 10 for the binary combination resistant vs not resistant (intermediate and sensitive), and in Supplementary Table 11 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 Table 9 - 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 10 - 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 11 - 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

1. Association of response with test classification

Supplementary Table 12 - Response by test classification in the development cohort

n (%)		resistant (N=41)	intermediate (N=43)	sensitive (N=32)
		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)	(22)

2. Multivariate Analysis of Development Cohort

Supplementary Table 13 - Multivariate Analysis of OS and PFS for 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 14 - Multivariate Analysis of OS and PFS for 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

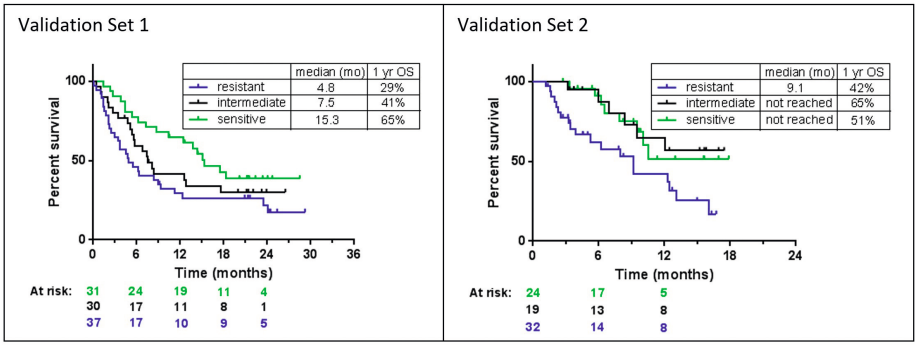
Supplementary Table 19 - PD-L1 status by test classification (pooled second line patients)

	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)

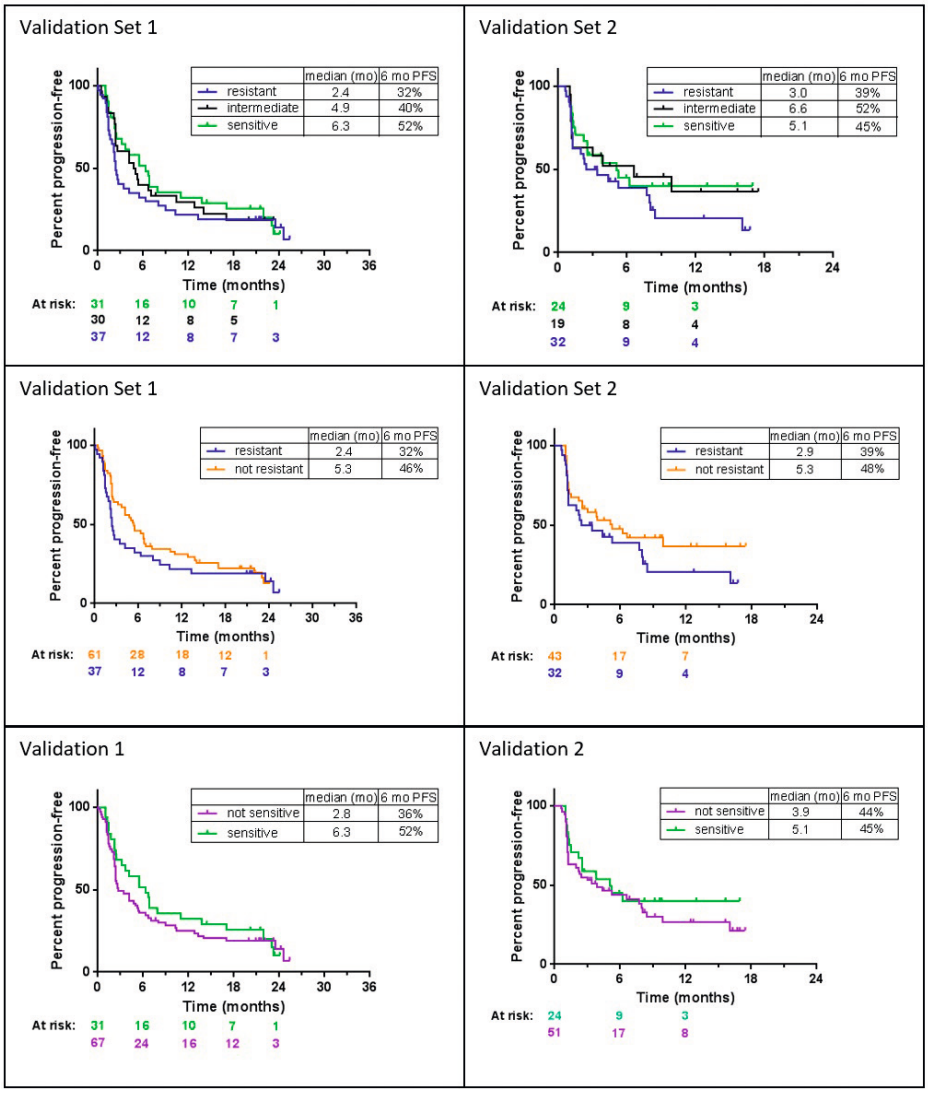
NA=not available

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

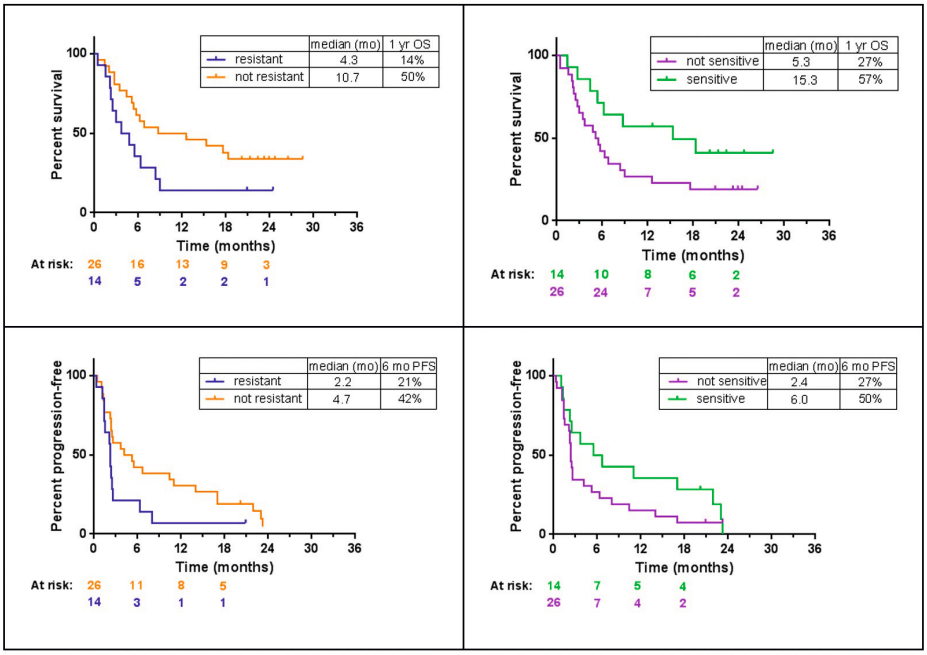
Biological Process	p value of association	FDR
Immune tolerance and suppression	0.0035	<0.1
Acute inflammatory response	0.0154	<0.2
Acute phase response	0.0170	<0.2
Cytokine production involved in immune response	0.0665	<0.4
Complement activation	0.1372	<0.5
Innate immune response	0.1523	<0.5
Angiogenesis	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
Type2 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
Type17 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



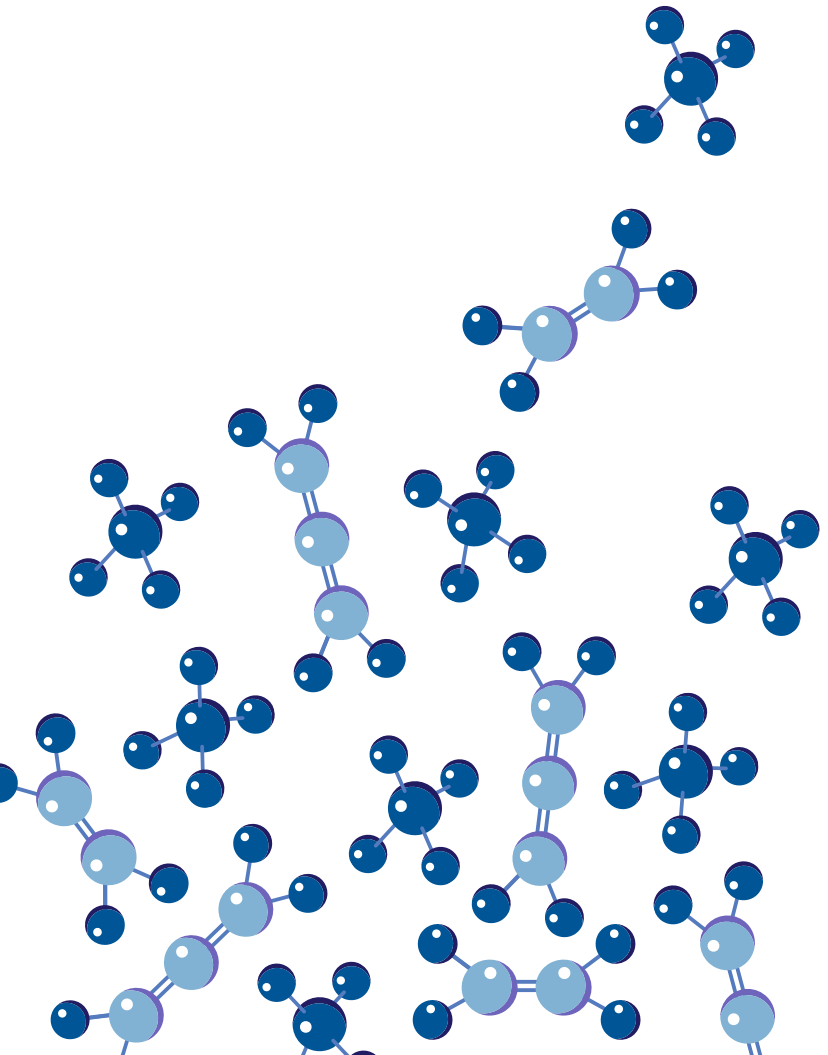
Supplementary Figure 5 - Kaplan-Meier plots of OS by test classification sensitive vs intermediate vs resistant in the Validation Sets



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.



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



5 Prediction of Response to Anti-PD-1 Therapy in Patients with Non-Small Cell Lung Cancer by Electronic Nose Analysis of Exhaled Breath

R. de Vries*, M. Muller*, V. van der Noort, W.S.M.E. Theelen, R.D. Schouten, K. Hummelink, S.H. Muller, M Wolf-Lansdorf, J.W.F. Dagelet, K. Monkhorst, A.H. Maitland – van der Zee, P. Baas, P.J. Sterk, M. M. van den Heuvel

Ann Oncol. 2019; 30(10): 1660-1666

*Contributed equally

Abstract

Background: Immune checkpoint inhibitors have improved survival outcome of advanced non-small-cell lung cancer (NSCLC). However, most patients do not benefit. Therefore, biomarkers are needed that accurately predict response. We hypothesized that molecular profiling of exhaled air may capture the inflammatory milieu related to the individual responsiveness to anti-programmed death ligand 1 (PD-1) therapy. This study aimed to determine the accuracy of exhaled breath analysis at baseline for assessing nonresponders versus responders to anti-PD-1 therapy in NSCLC patients.
Methods: This was a prospective observational study in patients receiving checkpoint inhibitor therapy using both a training and validation set of NSCLC patients. At baseline, breath profiles were collected in duplicate by a metal oxide semiconductor electronic nose (eNose) positioned at the rear end of a pneumotachograph. Patients received nivolumab or pembrolizumab of which the efficacy was assessed by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 at 3-month follow-up. Data analysis involved advanced signal-processing and statistics based on independent t-tests followed by linear discriminant and receiver operating characteristic (ROC) analysis.
Results: Exhaled breath data of 143 NSCLC patients (training: 92, validation: 51) were available at baseline. eNose sensors contributed significantly ($P < 0.05$) at baseline in differentiating between patients with different responses at 3 months of anti-PD-1 treatment. The eNose sensors were combined into a single biomarker with an ROC-area under the curve (AUC) of 0.89 [confidence interval (CI) 0.82–0.96]. This AUC was confirmed in the validation set: 0.85 (CI 0.75–0.96).
Conclusion: eNose assessment was effective in the noninvasive prediction of individual patient responses to immunotherapy. The predictive accuracy and efficacy of the eNose for discrimination of immunotherapy responder types were replicated in an independent validation set op patients. This finding can potentially avoid application of ineffective treatment in identified probable nonresponders.

Keywords
Immunotherapy; Responde prediction; Exhaled breath analysis; Electronic Nose; Lung cancer; Non-small cell lung cancer

Key message
Suspectibility to anti-programmed death ligand 1 (PD-1) therapy is reflected by a distinct exhaled breath profile. Therefore, the electronic nose qualifies as a noninvasice, point-of-care test outperforms the current standard for the selection of potential nonresponders to anti-PD-1 therapy.

1. Introduction

Immune checkpoint inhibitors (ICIs) have dramatically improved the treatment of advanced stage non-small-cell lung cancer (NSCLC) [183]. However, whereas profiling of the tumor has moved molecular-directed therapy toward highly accurate precision medicine for immunotherapy it is still in its infancy [40, 225]. Patient selection based on the expression of programmed death ligand 1 (PD-L1) as detected by immunohistochemistry (IHC) and more recently the tumor mutational burden can be used to improve the prediction of the likelihood of response to ICIs [40, 42, 226]. Despite its analytic and predictive limitations, PD-L1 expression testing by IHC presently remains the biomarker of choice to optimize clinical decision-making on treatment with ICIs [40].

If properly validated, molecular profiling of exhaled air may provide a noninvasive and rapid alternative for IHC assays. Exhaled breath contains thousands of volatile organic compounds (VOCs) that originate from both systemic and local metabolic processes. These can be associated with normal physiological processes, pathophysiological inflammation and proliferative or oxidative activity [227]. Electronic nose (eNose) technology can be applied for probabilistic pattern recognition of gas mixtures, by using multiple cross-reactive sensors that capture the complete mixture of VOCs in exhaled air without identification of the individual components [228, 229]. It has been shown that by using eNoses lung cancer [65, 230-232] and the epidermal growth factor receptor (EGFR) mutation [233] can be detected with relatively high accuracy values. In addition, we have recently shown that exhaled breath analysis allows clinical and inflammatory phenotyping of chronic airway diseases at the point-of-care and may therefore facilitate personalized anti-inflammatory strategies [234].

We hypothesized that exhaled breath analysis by eNose could provide identification of responders and nonresponders to anti PD-1 therapy in patients with advanced NSCLC. Associations between VOCs and response evaluation might occur because of a direct contribution of metabolite production by the tumor cells and/or the immunological/inflammatory host responses. This study aimed to first determine the accuracy of exhaled breath analysis at baseline for discrimination between responders and nonresponders to anti-PD-1 therapy in NSCLC patients using both a training and validation set. The second aim was to compare the predictive accuracy of the eNose with the PD-L1 biomarker for nonresponse to the anti-PD-1 therapy. By using this stepwise approach and transparent reporting, the study follows the recommendations of the STARD guidelines (Standards for Reporting of Diagnostic Accuracy Studies) [235] and the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) [236].

2. Methods

2.1 Study population

This was a prospective observational study linked to a real-world intervention in patients with advanced stage NSCLC who were eligible for treatment with anti-PD-1 therapy. Patients were included in the Netherlands Cancer Institute, Amsterdam, between March 2016 and February 2018. Patients who started treatment with pembrolizumab or nivolumab between March 2016 and April 2017 were included in the training set,

and patients who started treatment after April 2017 were included in the validation set. The patients were selected and received treatment in accordance with recent literature [183] and local guidelines. Patients who were scheduled to start treatment with pembrolizumab or nivolumab were asked to participate in this study. The ethics review board of the Netherlands Cancer Institute concluded in writing that Dutch legislation on human participation in research was not considered to be applicable, given the non-invasive nature of this study that merely added exhaled breath analysis to standard diagnostic procedures. Other clinical data (*e.g.* CT scan, blood tests and lung function) used in this study were collected for routine clinical practice and were subsequently handled by complying with the Dutch Personal Data Protection Act (WBP). Despite the waiver that was provided by the ethics review board, the purpose of adding the eNose to routine diagnostics was explained to the patients who all gave their oral consent. Details about the 'full eligibility criteria for treatment with immunotherapy' are presented in the supplementary material. Exclusion criteria for participating in this study were he recent (<12 hours) intake of alcohol or if patients were not willing or able to participate. In order to increase the applicability in clinical practice, there were no further restrictions.

2.2 Measurements

Within 2 weeks before the start of treatment, a computed tomography (CT) scan, blood tests and spirometry were carried out according to standard clinical care. The response to anti-PD-1 treatment was prospectively monitored using CT at 6-weekly intervals. Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 criteria were used for response evaluation and were reported accordingly: partial response (PR), stable disease (SD) and progressive disease (PD) [184]. Since our aim was to identify nonresponders to treatment with a high specificity, we conservatively grouped patients with PD as nonresponders and those with PR and SD as responders at 3-month follow-up.

2.2.1 Exhaled breath analysis

Exhaled breath analysis was carried out in duplicate with a 2-minute interval by eNose technology (SpiroNose) on the same day as the spirometry tests. The SpiroNose is a technically and clinically validated integration between routine spirometry and eNose technology, allowing stratification of patients in the doctor's office [65, 234]. All patients rinsed their mouth thoroughly 3 times with water. Subsequently, patients were asked to perform 5 tidal breaths followed by a single inspiratory capacity manoeuvre up to total lung capacity, a 5 second breath hold and slow (<0.4 L/S) expiration towards residual volume [65, 234]. Exhaled breath was real-time measured (<1 minute) by the SpiroNose, which is connected to an Ethernet cable for immediate secured data transmission to an online server for further automated analysis.

The SpiroNose has seven cross-reactive metal-oxide semiconductor sensors (sensors 1–7) for in duplicate sampling of exhaled air and the same set of sensors sampling ambient air for background correction (supplementary Figure S1 and Table S1). From each sensor signal, two variables are determined: (i) the highest sensor peak normalized to the most stable sensor, sensor 2, to minimize interarray differences and (ii) the ratio between the sensor peak and the breath hold (BH) point [234].

2.2.2 SpiroNose measurement setup

The eNose measurement setup used in this study included a mouthpiece, nose clamp, viral/bacterial filter (Lemon Medical GmbH) attached to a Masterscreen™ pulmonary function testing system (Masterscreen, Jaeger, CareFusion) and the SpiroNose (Figure S1, left) [65, 234]. The SpiroNose consists of 8 separate sensor arrays, 4 reference sensor arrays to monitor environmental air and 4 sensor arrays used to monitor the VOCs in exhaled breath (Figure S1, right). In total, the SpiroNose contains 7 different metal oxide semiconductor sensors (Table S1) and each sensor is present in duplicate in both the reference and breath-monitoring sensor arrays. The MOS sensor stability was verified, as previously described, using the standard test gas for pulmonary diffusion capacity measurements as quality control gas every morning before patient measurements [65, 234]. The SpiroNose provides a simple approach to exhaled breath analysis which has shown to be able to discriminate between asthma, COPD, lung cancer and healthy controls [65] and to phenotype chronic airway disease [234].

2.2.3 PD-L1 IHC

The PD-L1 IHC 22C3 pharmDx assay (Agilent Technologies, Santa Clara, CA) was used to assess PD-L1 expression in the baseline biopsies. For 34 patients we used the 22C3 antibody kit, done in our hospital. Immunohistochemistry of the formalin-fixed paraffin-embedded (FFPE) tumor samples was performed on a BenchMark Ultra autostainer (Ventana Medical Systems). Briefly, paraffin sections were cut at 3 µm, heated 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 48 minutes at 950C PD-L1 clone 22C3 (DAKO) was detected using 1/40 dilution, 1 hour at RT. 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 two patients, the PD-L1 score with the 22C3 was done in another hospital. For 3 other patients the SP263 immunohistochemistry was used (from whom 2 had a score >50% and one <1%) and for one patient the antibody was unknown (score >50%). Since these PD-L1 scores were used for receiving pembrolizumab in first line (except for the patient with a score <1%), we also used these in our analysis.

2.3 Data processing

The processing of the eNose sensor signals including filtering, de-trending, ambient correction and peak detection was automatically carried out by the standard eNose software as was previously published [65, 234]. The signal processing resulted in a .csv file containing the selected parameters (sensor peak- and peak/BH ratios) serving as the source document for statistical analysis.

2.3.1 Statistical analysis

SPSS (IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.) and MatLab (2017B, MathWorks, Natick, MA) were used for data analysis. Descriptive statistics were expressed as mean ± SD if data were normally distributed and as median (interquartile range) if data were non-normally distributed. Between-group

comparisons were carried out using Mann–Whitney U tests, two-sample unpaired t-tests or chi-square tests.

2.3.2 Sample size calculation

The sample size calculation of the *training set* was based on results of an independent pilot study [237] using a previous version of the SpiroNose for discrimination between responders and non-responders to immunotherapy. In this pilot we found that the largest difference between responders and non-responders among the values of the individual sensors was 0.9 times the common standard deviation of the groups for that sensor. In our *training set* (where we would combine the data from different sensors into one marker) we expected a difference of at least this size. We then wanted to include enough patients so that a difference of this size equal at least four times the standard error of the groups for our new marker. This would require the inclusion of at least 20 patients per group. To compensate for a possible skewed distribution or unexpected differences between the pilot study and the *training set* we decided to double the total number of patients and recruit a total of 80 patients, which we believed would yield at least 20 patients per group even if response is not yet known at the time of inclusion.

The sample size estimation for the *validation set* was based on our aim to compare the area under the receiver operating characteristic curve (ROC-AUC) of the eNose and the current biomarker PD-L1 for discrimination between responders and non-responders to anti-PD-1 treatment in NSCLC patients. PD-L1 has an AUC of approximately 0.70 [238]. We considered a new biomarker to be successful if the two-sided 95% DeLong confidence interval of the AUC is >0.70. In our *training set* we found an AUC of 0.89 for discrimination between responders and non-responders to immunotherapy. Using a 20000 bootstrapped sample from our *training cohort*, we established that the power of showing this biomarker to outperform PD-L1 IHC in a *validation cohort* of 51 patients is.

2.3.3 Exhaled breath analysis

All patients rinsed their mouth thoroughly 3 times with water. Subsequently, patients were asked to perform 5 tidal breaths followed by a single inspiratory capacity manoeuvre up to total lung capacity, a 5 second breath hold and slow (<0.4 L/S) expiration towards residual volume [65, 239] . Exhaled breath was real-time measured (<1 minute) by the SpiroNose, which is connected to an Ethernet cable for immediate secured data transmission to an online server for further automated analysis. The normalized sensor peaks and peak/BH ratios were compared between groups by means of independent sample t-tests. The variables that discriminated ($P < 0.05$) between responders and nonresponders were selected for further analysis. The t-tests were internally validated by 1000 iterations of bootstrap. Subsequently, linear discriminant analysis was carried out using the selected variables. A discriminant function was calculated that best distinguished between responders and nonresponders. The accuracy of this model was defined as the percentage correctly classified patients in the training set. Crossvalidation using the leave-one-out method was used to calculate the cross-validated accuracy value (CVV, %). The discriminant scores were used to construct receiver operating characteristics (ROC) curves. To test external validity, the discriminant function obtained from the training set was examined in the independent validation set and compared

based on the area under the curves (AUC) of the ROC curves. Discriminant scores were converted into predicted probabilities of nonresponse to aid interpretation. Details regarding the computation are provided in the supplementary material. Finally, in the validation set, we identified a cut-off point on the predicted probabilities scale to identify nonresponders to anti-PD-1 therapy with 100% specificity and to calculate the sensitivity, positive predicted value (PPV) and negative predictive value (NPV) for this cut-off point. To provide additional illustration of our results, a survival analysis with a Kaplan–Meier curve was carried out.

2.3.3 PD-L1 IHC

The PD-L1 expression of the baseline biopsies was used to (i) carry out an independent t-test to assess the discrimination between responders and nonresponders, (ii) construct an ROC curve predictive for the selection of nonresponders and (iii) create a table of frequencies to report the classification results where the PD-L1 IHC was considered positive when the tumor-proportion score was ≥1%.

3. Results

In total, 143 NSCLC patients were included in this study of which 92 patients were used for *training* and 51 patients for *validation* of the results (Table 1). In the *training set*, among the baseline characteristics only the Eastern Cooperative Oncology Group (ECOG) performance status showed a significant difference ($p<0.001$) between responders (n=42) and non-responders (n=50). In the *validation set* there was a significant difference ($p=0.03$) in the number of patients with a KRAS mutation between responders (n=23) and non-responders (n=28). There were no significant differences in any other baseline characteristic between responders and non-responders in both *training* and *validation set*. In addition, the baseline characteristics of patients in the *training set* and *validation set* significantly differed regarding choice of treatment ($p<0.001$), line of treatment ($p<0.001$), history of treatment ($p<0.001$), PD-L1 scores ($p<0.001$) and lung function ($p=0.03$) (Table 1).

Table 1. Demograpic and disease characteristics of the patients at baseline.

	Training		Validation	
	Non-responders (N=50)	**Responders (N=42)**	**Non-responders (N=28)**	**Responders (N=23)**
Patient				
Male sex – no.(%)	22 (44)	20 (48)	15 (54)	11 (48)
Age (years) – mean (±SD)	63.6 (10.0)	63.3 (8.5)	63.1 (8.1)	63.5 (10.9)
BMI – mean (±SD)	24.4 (4.3)	25.3 (5.4)	26.0 (5.8)	26.1 (4.4)
Non-smokers – no.(%)	5 (10)	2 (5)	2 (7)	0 (0)
Pack years – median (IQR)	30 (17)	33 (25)	43 (24)	36 (16)
Caucasian – no.(%)	47 (94)	40 (95)	26 (93)	22 (97)
Performance score ≥ 2 – no.(%)[1]	9 (18)	3 (7)	9 (32)[1]	0 (0) [1]
Tumor characteristics				
Pathology – no. (%)				
Adenocarcinoma	36 (72)	25 (60)	17 (61)	19 (83)
Squamous	6 (12)	13 (30)	4 (14)	2 (9)
Other	8 (16)	4 (10)	7 (25)	2 (9)
Mutations – no.(%)				
EGFR positive	1 (2)	1 (2)	2 (7)	1 (4)
KRAS positive	20 (40)	13 (31)	8 (29)[2]	13 (57) [2]
PD-L1 – no.(%)[3]				
Negative	6 (32)	7 (54)	12 (50)	3 (19)
Positive >1%	13 (68)	6 (46)	12 (50)	13 (81)
Positive>50%	0 (0)	3 (23)	11 (46)	13 (81)
Brain metastases – no.(%)	10 (20)	9 (21)	10 (36)	3 (13)
Treatment				
Current – no.(%)				
Nivolumab[4]	47 (94)	41 (98)	16 (57)	8 (35)
Pembrolizumab[4]	3 (6)	1 (2)	12 (43)	15 (65)
Line of treatment – no.(%)				
1st line[4]	1 (2)	0 (0)	6 (21)	9 (39)
≥ 2 line[4]	49 (98)	42 (100)	22 (79)	14 (61)
History – no.(%)				
Previous RT	29 (58)	28 (67)	21 (75)	14 (61)
Thoracic RT	18 (36)	18 (43)	15 (54)	7 (30)
Platinum doublet[4]	3 (76)	27 (64)	16 (57)	10 (43)
CCRT[4]	9 (18)	12 (29)	5 (18)	4 (17)
Platinum doublet and CCRT[4]	1 (2)	2 (5)	0 (0)	0 (0)
Also had a TKI[4]	1 (2)	1 (2)	1 (4)	0 (0)
None[4]	1 (2)	0 (0)	6 (21)	9 (39)
Comorbidities				
COPD – no.(%)	29 (54)	25 (57)	14 (50)	14 (61)
GOLD I	9 (18)	5 (12)	3(11)	3 (13)
GOLD II	16 (32)	15 (36)	6 (21)	4 (17)
GOLD III	4 (8)	5 (12)	5 (18)	7 (30)
FEV1 – mean (±SD)[5]	73.0 (18.6)	72.5 (20.7)	62.7 (22.8)	68.0 (20.8)
Auto-immune disease - no.(%)	3 (6)	4 (10)	6 (21)	2 (9)

Table 1. Demograpic and disease characteristics of the patients at baseline. *(Continued)*

	Training		Validation	
	Non-responders (N=50)	**Responders (N=42)**	**Non-responders (N=28)**	**Responders (N=23)**
COPD treatment no.(%)				
Short-acting bronchodilator	9(18)	8(19)	6 (21)	4 (17)
Long-acting bronchodilator	14 (28)	16(38)	5 (18)	5 (22)
Oxygen supplement	1(2)	2 (5)	0	0

Footnotes: *[1] Significant difference, P = 0.003; [2] Significant difference, P = 0.03; [3]Not abailable for all patients, percentage shown in percentage of known cases; [4] Significant difference between training and validation set, P<0.001; [5] Significant difference between training and validation set, P=0.033;*

Abbreviations: *SD, Standard deviation; BMI, Body mass index; IQR, interquartile range; FEV1, Forced expiratory volume in 1 second; 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 symptom, 1 indicates mild symptoms and above 1 indicates greater disability; PD-L1, Programmed death ligand 1; EGFR, epidermal growth factor receptor; KRAS, Kirsten Rat Sarcoma viral oncogene; RT, Radiation therapy; Platinum doublet, Patients who received platinum doublet therapy as previous line of treatment, dependent on their diagnosis; CCRT, Concurrent chemo radiation therapy; Platinum and CCRT, Patients who received both platinum doublet and chemo radiation therapy in a short time window; TKI, Tyrosine kinase inhibitor; COPD, Chronic obstructive pulmonary disorder; GOLD, Global Initiative for chronic obstructive lung disease*

3.1 Exhaled breath analysis:

3.1.1 Training set

Independent t-test analysis showed that sensor 3 (p<0.001), sensor 5 (p=0.01), sensor 1_BH (p=0.04) and sensor 5_BH (p=0.001) were significantly different at baseline between responders and non-responders to anti-PD-1 therapy. Subsequent discriminant analysis showed a cross-validated accuracy value of 82%. The ROC-AUC ± 95% confidence interval (CI) after internal cross-validation reached 0.89 (CI: 0.82 - 0.96) (Figure 1A and 1B).

3.1.2 The equation

The equation below provides the mathematical model resulting from the linear discriminant analysis of the *training set* and reflects the predicted probability of non-response for each patient in terms of his/her eNose assessment. This equation was used to determine the predicted probabilities of non-response for the patients in the *validation set* and can be used to obtain predictions for future patients. Labels S3 and S5 indicate the normalized sensor peak of sensor 3 and sensor 5, respectively. $S1_{BH}$ and $S5_{BH}$ symbolize the ratio between the highest sensor peak and the BH point for sensor 1 and sensor 5, respectively. Additional information about the mathematical model is provided in the *Online Supplement.*

$$\text{Probability of PD} = \frac{1}{1+ e^{29.090865*\text{S3} - 3.222506*\text{S5} + 0.918908*\text{S1}_{\text{BH}} - 4.976302*\text{S5}_{\text{BH}} - 26.843187}}$$

3.1.3 Validation set

The ability to predict response to anti-PD-1 therapy from exhaled breath, with sensor 3, sensor 5, sensor 1_BH and sensor 5_BH as input for the model obtained from the *training set*, was confirmed in the independent *validation set*. Breath profiles of responders were distinguished from non-responders to anti-PD-1 therapy with a ROC-AUC of 0.85 (CI: 0.7-0.96) (Figure 1C and 1D). To exclude erroneous withholding of anti-PD-1 therapy

based on the model, we selected a cut-off value of 0.72 for the predicted probabilities of membership in the non-responder group in order to obtain 100% specificity and thereby 0% false positive prediction for non-response to anti-PD-1 (Figure 2). This cut-off value resulted in a sensitivity of 43%, a PPV of 100% and a NPV of 59% for non-response to anti-PD-1 therapy *(*Supplementary Table S3A*).* 12 out of the 51 (24%) NSCLC patients showed a predicted probability of membership in the non-responder group ≥0.72. These 12 NSCLC patients were all non-responders to anti-PD-1 therapy (Figure 2). The predicted probabilities of non-response as well as the response evaluation of all 51 patients in the *validation set* are provided in Supplementary Table S2*,* allowing computations of the sensitivity, specificity, PPV and NPV resulting from different cut-off points. Finally, both the progression free survival and overall survival curves are provided in Supplementary Figure S3*.*

3.1.4 PD-L1 immunohistochemistry in the validation set

PD-L1 expression was available in 40 out of the 51 NSCLC patients. Independent t-test analysis showed that the PD-L1 expression was significantly different (p=0.03) at baseline between responders and non-responders to anti-PD-1 therapy. The PD-L1 expression of responders was distinguished from non-responders with a ROC-AUC of 0.66 (CI: 0.49-0.83) (Figure 1E). The identification of non-responders resulted in a specificity of 81%, a sensitivity of 50%, a PPV of 80%, and a NPV of 52% *(*Supplementary Table S3B).

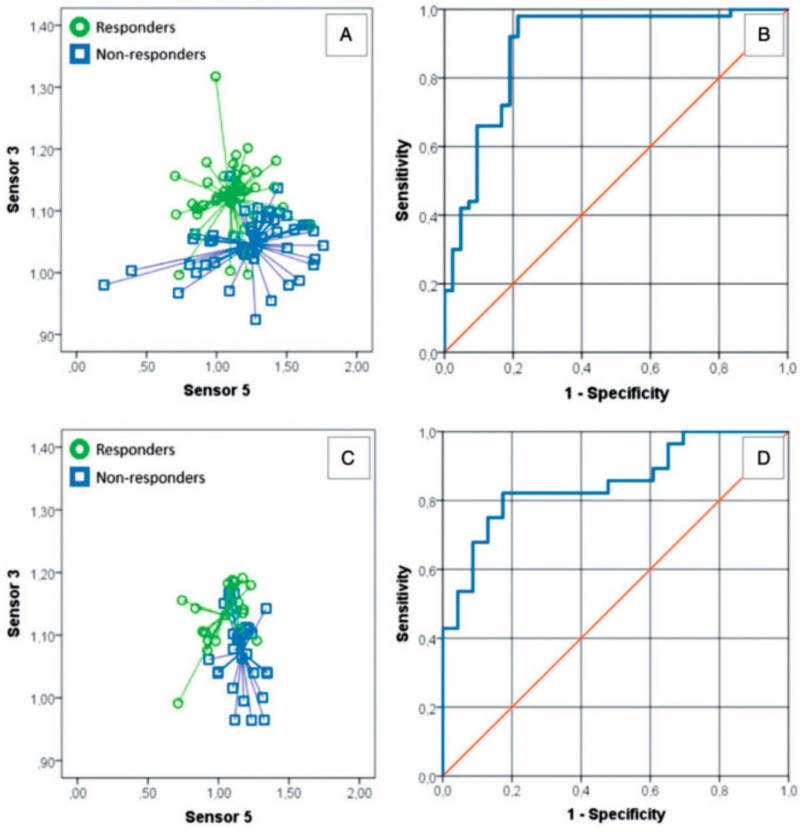


Figure 1 - Results

Notes: *(A) Training set: Two-dimensional plot showing the discrimination of breath profiles between non-small-cell lung cancer (NSCLC) patients who are clinical responders [partial response (PR), stable disease (SD)] and nonresponders (progressive disease [PD]) to immunotherapy at baseline. The x and y axes represent normalized sensor values. (B) Receiver operating characteristic (ROC) curve with line of identity of the breath profile. Discriminant function (representing sensor 3, sensor 5, sensor 1_BH and sensor 5_BH) predictive for the selection of nonresponders [area under the curve (AUC): 0.89]. (C) Validation set: Two-dimensional plot showing the discrimination of breath profiles between NSCLC patients that are clinical responders (PR and SD) and nonresponders (PD) to immunotherapy at baseline. (D) ROC curve with line of identity of the breath profile discriminant function, obtained from the training set and predictive for the selection of non-responders (AUC: 0.85). (E) ROC curve with line of identity of the PD-L1 immunohistochemistry predictive for the selection of non-responders to anti-PD-1 therapy (AUC: 0.66).*

Discussion

This study shows that exhaled breath analysis by eNose allows discrimination at baseline between responders and non-responders to anti-PD-1 treatment in both a *training* and *validation set*. This was substantially better than obtained with the current clinical standard biomarker PD-L1. The results of the *validation set* showed that ineffective anti-PD-1 therapy could potentially be saved in 24% of the NSCLC patients without erroneously withholding anyone effective treatment. These results indicate that susceptibility to anti-PD-1 therapy is reflected by a distinct exhaled molecular fingerprint. To our knowledge, this is the first study that applied an eNose to identify responsiveness to anti-PD-1 therapy. Our study extends the work of Shlomi et al. [233], who were able to discriminate between patients with lung cancer that harbor the EGFR mutation from those with wild-type EGFR using eNose technology. The assessment of the mutation status has become essential for treatment decision making in advanced stage NSCLC and has changed the treatment strategy of NSCLC to individualized therapy [240]. Given the clinically applicable technology, the present data may qualify breath assessment as a real-time tool for stratification of patients with NSCLC. This meets the demands of modern medicine, in which treatment decisions are taken based on the patient's individual subtype rather than a classical diagnosis [240]. Recently, studies have been published showing that different cancer cell lines produce different VOCs [241, 242], making it biologically plausible that these different cell lines also lead to different VOC profiles that can be measured in vivo using an eNose. Interestingly, one of the cross-reactive eNose sensors, sensor 3, which has the highest sensitivity to methane and natural gas, has a major influence on the eNose data in this study. It is expected that this sensor is also sensitive to other hydrocarbons, which appear to be different between cell lines of various origins [241, 242].

Broadly speaking, there are two approaches to exhaled breath analysis [229]. Methods based on mass spectrometry (MS) aim to detect individual VOCs and in some cases identify them, which is particularly useful for pathophysiologic research. However, identification of individual molecular compounds using MS is not possible or suitable yet for clinical implementation. On the other hand, eNoses are based on cross-reactive nonspecific sensor arrays. VOCs competitively interact with the sensors allowing multiple VOCs to bind to the same sensor based on their affinity for both the sensor and its substrate [228]. Likewise, multiple sensors interact with the same volatile. This is comparable to the powerful mammalian olfactory system [243] and results in a pattern of firing sensors that is driven by the complete mixture of VOCs. Such pattern recognition allows probabilistic analyses which results in the present ROC curves that are suitable for clinical diagnostics. However, without additional information on the biochemical pathways that contribute to the production of the measured VOCs, it remains to be established whether the associations between VOCs and response evaluation are a direct effect of metabolite production by the tumor cells and/or the immunological/inflammatory host responses. This does not hamper the clinical application of a breath test for prediction of response to anti-PD-1 therapy.

An ideal biomarker is minimally invasive, easy to collect, reliable, inexpensive and can be used to accurately identify a treatment response phenotype, to measure changes in disease activity or to confirm a diagnosis. This study used an eNose to carry out exhaled breath analysis as part of routine assessment at the point-of-care. In comparison with the currently used biomarker the eNose outperforms PD-L1 IHC in all areas. The strength of our study is the use of an independent validation set that confirmed the results of the training set. In this study, we chose to combine patients receiving nivolumab or pembrolizumab because both treatments are interchangeable since they both interfere with the interaction between PD-1 and PDL1, whose unimpeded interaction downregulates T cells, allowing cancer cells to evade immune surveillance [244].

A potential limitation of the present study is represented by its design, in which anti-PD-1 therapy was given in the clinical care setting rather than a randomized controlled trial. Therefore, no comparison has been made to other biomarkers, such as tumor mutational burden [245], as these biomarkers are not yet clinically validated for the selection of treatment [246]. One could argue that this real-world design brings the evaluation of biomarkers closest to daily practice. However, it cannot be excluded that this has introduced selection-bias, for instance based on considering PD-L1 IHC in clinical treatment decisions. In an effort to eliminate the chance of unjustifiable withholding of anti-PD-1 therapy based on the eNose results, we concentrated on the identification of nonresponders to treatment. However, earlier biomarker research in this patient group, notably on PD-L1 expression, focused on identifying patients who will benefit most from antiPD-1 therapy [39, 40]. Therefore, we provide all necessary information (supplementary Table S2) for the reader to compute the discriminative power between patients with PR, SD and PD for both PD-L1 IHC and the predicted probabilities obtained from the eNose results.

In conclusion, before starting anti-PD-1 therapy, responders and nonresponders can be distinguished by eNose analysis of exhaled breath in patients with NSCLC. These results were validated in an independent set of patients with advanced stage NSCLC. The equations for calculating individual patient probabilities of responsiveness to anti-PD-1 therapy are provided for future clinical application. These results show that breath analysis by eNose can potentially avoid application of ineffective treatment in identified probable nonresponders. This way individual patients might be saved from unnecessary delays and start treatment with a better alternative. The present results merit taking the next step, namely, a large prospective multi-center study on NSCLC outcome.

Supplemental material

The full supplemental material:

Content included in this thesis:
Supplemental methods
Supplemental results

Supplemental methods

Below are the criteria that apply in the Netherlands to qualify for the use of nivolumab/ pembrolizumab.

Nivolumab
Compassionate use program
In august 2015 the compassionate use program started, which meant nivolumab was provided by Bristol-Myers Squibb (BMS) in eight different hospitals in the Netherlands (NCT02475382). More information about the eligibility criteria and protocol have previously been published by Schouten et al. [24].

Standard of care
In March 2016 nivolumab became available as routine clinical care for patients with advanced squamous non-small cell lung cancer, non-squamous followed soon thereafter. The inclusion criteria were less strict, however, patients still had received at least one previous line of anticancer treatment, results of blood tests consistent with adequate organ functions, and a good clinical performance (PS0/1, according WHO [247]). These patients are already described in the publication from Schouten et al [24].

Pembrolizumab
Second line
Patients who were treated with pembrolizumab in second line, were participating in a trial with pembrolizumab (NCT02492568). This was a randomized phase II, 2-arm study of pembrolizumab after high dose radiation (SBRT) versus pembrolizumab alone in patients with advanced non-small cell lung cancer). Only patients who did not receive SBRT were included in the current eNose study.

First line
Patients who were treated with pembrolizumab in the first line, were treated according to standard of care. They all had a PD-L1 score of >50% at baseline.

SpiroNose measurement setup
The eNose measurement setup used in this study included a mouthpiece, nose clamp, viral/bacterial filter (Lemon Medical GmbH) attached to a MasterscreenTM pulmonary function testing system (Masterscreen, Jaeger, CareFusion) and the SpiroNose (Figure S1, left) [65, 239]. The SpiroNose consists of 8 separate sensor arrays, 4 reference sensor arrays to monitor environmental air and 4 sensor arrays used to monitor the VOCs in exhaled breath (Figure S1, right). In total, the SpiroNose contains 7 different metal oxide semiconductor sensors (Table S1) and each sensor is present in duplicate in both the reference and breath-monitoring sensor arrays. The MOS sensor stability was verified, as previously described, using the standard test gas for pulmonary diffusion capacity measurements as quality control gas every morning before patient measurements [65, 239]. The SpiroNose provides a simple approach to exhaled breath analysis which has shown to be able to discriminate between asthma, COPD, lung cancer and healthy controls [65] and to phenotype chronic airway disease [239].

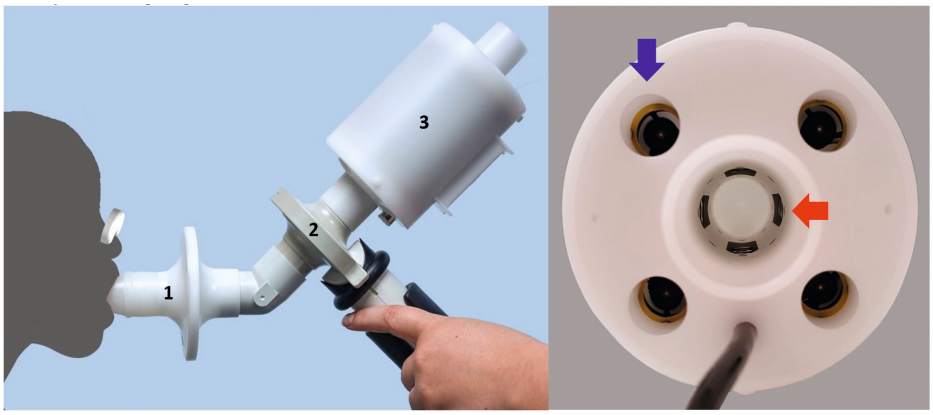


Figure S1 - SpiroNose measurement setup
Left: (1) Mouthpiece, nose clamp and bacteria filter, (2) Spirometer, (3) SpiroNose. Right: Front view of the SpiroNose and the positioning of the sensor arrays. Red arrow: 4 sensor arrays monitoring exhaled breath. Blue arrow: 4 reference sensor arrays monitoring ambient VOCs.

Table S1 - Sensors of the SpiroNose

	Type	Highest sensitivity for:	Range (ppm)
Sensor 1	TGS 2602	VOCs (e.g. toluene) and odorous gases (e.g. ammonia and hydrogen sulphide)	1 - 30
Sensor 2	TGS 2610	butane and propane	500 - 10.000
Sensor 3	TGS 2611-COO	methane and natural gas	500 - 10.000
Sensor 4	TGS 2600	air contaminants (e.g. hydrogen, carbon monoxide and ethanol)	1 - 30
Sensor 5	TGS 2603	air contaminants (e.g. trimethylamine, methyl mercaptan)	1 - 30
Sensor 6	TGS 2620	alcohol and solvent vapors	50 - 5.000
Sensor 7	TGS 2612	methane, propane and iso-butane	500 - 10.000

ppm: parts per million, VOCs: volatile organic compounds

Supplemental results

Results: Validation set

Table S2 - Response evaluation for patients in the validation set.

ID	Treatment	Line	TRUE response	RECIST response	PD-L1 TEST	PD-L1 score	Probability score	Cut-Off 0.5	Cut-Off 0.72
14	Pembrolizumab	1	Non-responder	PD	Positive	50	0,987758	Non-responder	Non-responder
45	Nivolumab	2	Non-responder	PD	Negative	0	0,986647	Non-responder	Non-responder
30	Pembrolizumab	1	Non-responder	PD	Positive	75	0,973996	Non-responder	Non-responder
5	Nivolumab	3	Non-responder	PD	Negative	0	0,953076	Non-responder	Non-responder
2	Nivolumab	2	Non-responder	PD	Negative	0	0,949185	Non-responder	Non-responder
46	Pembrolizumab	2	Non-responder	PD	Positive	90	0,889516	Non-responder	Non-responder
38	Pembrolizumab	2	Non-responder	PD	Positive	60	0,889284	Non-responder	Non-responder
50	Pembrolizumab	3	Non-responder	PD	Positive	100	0,865832	Non-responder	Non-responder
27	Nivolumab	2	Non-responder	PD	Negative	0	0,837278	Non-responder	Non-responder
19	Nivolumab	2	Non-responder	PD	Negative	0	0,811239	Non-responder	Non-responder
11	Nivolumab	2	Non-responder	PD			0,73746	Non-responder	Non-responder
12	Nivolumab	2	Non-responder	PD			0,727406	Non-responder	Non-responder
10	Pembrolizumab	1	Responder	SD			0,716845	Non-responder	Responder
17	Nivolumab	2	Non-responder	PD	Negative	0	0,614834	Non-responder	Responder
33	Nivolumab	2	Non-responder	PD	Negative	0	0,594958	Non-responder	Responder
21	Pembrolizumab	2	Non-responder	PD	Positive	70	0,572296	Non-responder	Responder
25	Pembrolizumab	1	Responder	SD	Positive	50	0,496864	Responder	Responder
48	Pembrolizumab	2	Non-responder	PD	Positive	5	0,483539	Responder	Responder
26	Nivolumab	2	Non-responder	PD			0,475168	Responder	Responder
35	Pembrolizumab	1	Non-responder	PD	Positive	100	0,423106	Responder	Responder
18	Pembrolizumab	1	Non-responder	PD	Positive	100	0,421781	Responder	Responder
37	Pembrolizumab	1	Responder	PR	Positive	95	0,403842	Responder	Responder
3	Pembrolizumab	2	Non-responder	PD			0,384225	Responder	Responder
15	Nivolumab	2	Non-responder	PD	Positive	50	0,333844	Responder	Responder
29	Pembrolizumab	2	Responder	PR	Positive	90	0,330109	Responder	Responder
9	Nivolumab	2	Non-responder	PD	Negative	0	0,31903	Responder	Responder

Table S2 - Response evaluation for patients in the validation set. *(Continued)*

ID	Treatment	Line	TRUE response	RECIST response	PD-L1 TEST	PD-L1 score	Probability score	Cut-Off 0.5	Cut-Off 0.72
23	Nivolumab	2	Non-responder	PD	Negative	0	0,311778	Responder	Responder
43	Pembrolizumab	1	Responder	PR	Positive	70	0,294397	Responder	Responder
34	Pembrolizumab	2	Responder	PR	Positive	70	0,258442	Responder	Responder
13	Nivolumab	2	Responder	SD	Positive	75	0,190093	Responder	Responder
49	Pembrolizumab	2	Responder	SD	Positive	80	0,186393	Responder	Responder
51	Pembrolizumab	1	Responder	PR	Positive	100	0,156492	Responder	Responder
28	Pembrolizumab	2	Responder	PR	Negative	0	0,148911	Responder	Responder
40	Pembrolizumab	1	Responder	SD	Positive	50	0,136329	Responder	Responder
36	Nivolumab	2	Non-responder	PD	Negative	0	0,110182	Responder	Responder
41	Nivolumab	2	Responder	SD			0,108573	Responder	Responder
16	Pembrolizumab	1	Responder	SD			0,106446	Responder	Responder
24	Pembrolizumab	1	Responder	SD	Positive	100	0,104979	Responder	Responder
42	Pembrolizumab	1	Non-responder	PD	Positive	99	0,088359	Responder	Responder
22	Pembrolizumab	1	Responder	PR	Positive	50	0,074115	Responder	Responder
47	Pembrolizumab	1	Non-responder	PD	Positive	70	0,071955	Responder	Responder
7	Nivolumab	2	Non-responder	PD	Negative	0	0,071068	Responder	Responder
32	Pembrolizumab	2	Responder	PR	Negative	0	0,05683	Responder	Responder
8	Nivolumab	2	Non-responder	PD	Negative	0	0,053161	Responder	Responder
39	Pembrolizumab	2	Responder	SD	Positive	70	0,050501	Responder	Responder
44	Nivolumab	2	Responder	SD			0,038403	Responder	Responder
31	Nivolumab	2	Responder	SD			0,037547	Responder	Responder
6	Nivolumab	2	Responder	SD			0,036347	Responder	Responder
1	Nivolumab	2	Responder	SD			0,035672	Responder	Responder
20	Nivolumab	2	Responder	SD	Negative	0	0,029033	Responder	Responder
4	Nivolumab	2	Responder	PR	Positive	90	0,025823	Responder	Responder

Response characteristics per patient for the RECIST criteria, PD-L1 criteria and the test with the eNose. Also, the line of treatment is revealed. The response evaluation obtained by the gold standard are provided in the columns "Response group" and "RECIST response". The PD-L1 IHC results are shown in the columns "PD-L1 IHC" and "PD-L1 score". The eNose results are provided in the column "eNose probability score for non-response". The response classification based on the eNose results are provided for two cut-off points (0·50 and 0·72). IHC: Immunohistochemistry, PD-L1: Programmed death ligand 1, PD: progressive disease, SD: stable disease, PR: partial response, - represents missing data.

Table S3 - Validation set: Table of frequencies for the classification of response based on (A) eNose and (B) PD-L1 immunohistochemistry.

	Gold standard: RECIST 1.1		
eNose	Non-responder (PD)	Responder (PR and SD)	Total
Predicted probability ≥ 0.72	12	0	12
Predicted probability < 0.72	16	23	39
Total	25	23	51

	Gold standard: RECIST 1.1		
PD-L1 immunohistochemistry	Non-responder (PD)	Responder (PR and SD)	Total
PD-L1 < 1% (negative)	12	3	15
PD-L1 ≥ 1% (positive)	12	13	25
Total	24	16	40

Table S4 - Multiple cut-off points to compare PD-L1 and eNose with regard to the prediction of non-response to anti-PD-1 therapy AUC: Area Under the Curve, PPV: Positive Predicted Value, NPV: Negative Predicted Value

Biomarker	AUC	Cut-off point	Sensitivity	Specificity	PPV	NPV
PD-L1	0.66 CI: 0.49 - 0.83	<1% vs ≥1%	50%	81%	80%	52%
PD-L1		<50% vs ≥50%	54%	81%	81%	54%
eNose	0.85 CI: 0.75 - 0.96	<0.72 vs ≥0.72	43%	100%	100%	59%
eNose		<0.5 vs ≥0.5	54%	96%	94%	63%
eNose		<0.32 vs ≥0.32	75%	83%	84%	73%

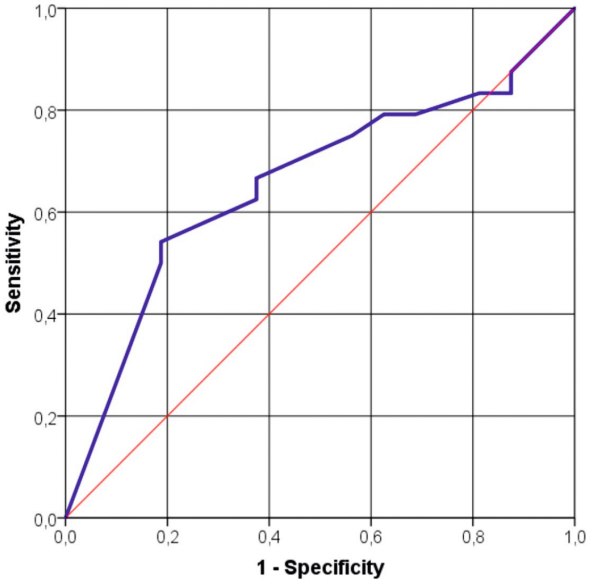


Figure S3 - Validation set: ROC curve with line of identity of the PD-L1 immunohistochemistry predictive for the selection of non-responders to anti-PD-1 therapy (AUC: 0.66). PD: progressive disease, PR: partial response, SD: stable disease, RECIST: Response Evaluation Criteria in Solid Tumors, PD-L1: programmed death ligand 1

The only way PD-L1 could reach 100% specificity in our *validation cohort* is by declaring every patient a prospective responder, hence obtaining 0% sensitivity (Table S2). Conversely, with the eNose, a specificity of 83% (the value available in our cohort closest to the 81% achieved by PD-L1) is obtained by setting the cut-off point to a predicted probability of non-response of 0.32. At this cut-off point the eNose reached a sensitivity of 75%, a PPV of 84% and a NPV of 73% (Table S3).

Results: Survival analysis

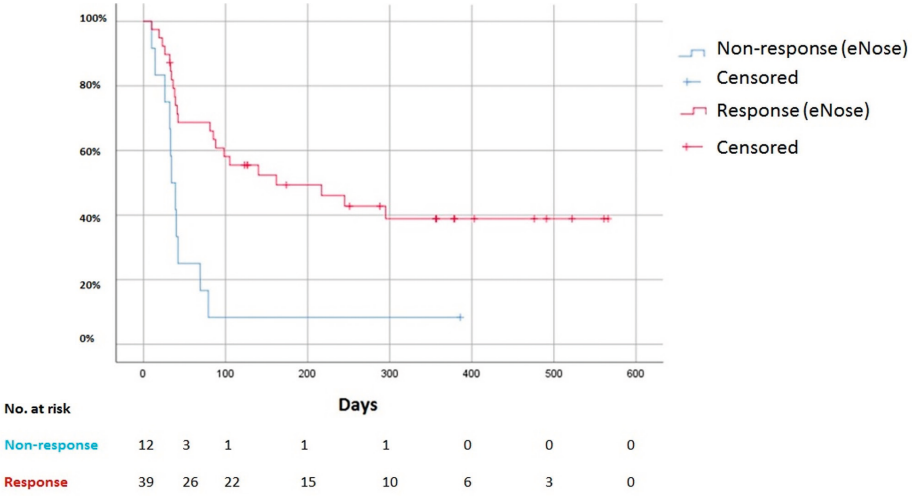


Figure S4A – Progression Free Survival: In this figure the progression free survival (PFS) analysis of the two groups based on the eNose test (with a cut-off of 0.72) of the validation set is shown with a Kaplan Meier curve. In the non-responder group according to the eNose test, one patient is censored (386 days). In the responder group according to the eNose test, 9 patients were censored in the first year of follow-up. The median progression free survival was 34 days (IQR: 26 - 42 days) in the non-responder group and 162 days (IQR: 39 - not reached (NR) days) in the responder group. With a log rank (Mantel-Cox) test, the survival curve showed a significant difference between responders and non-responders according to their group based on the eNose (p=0.001).

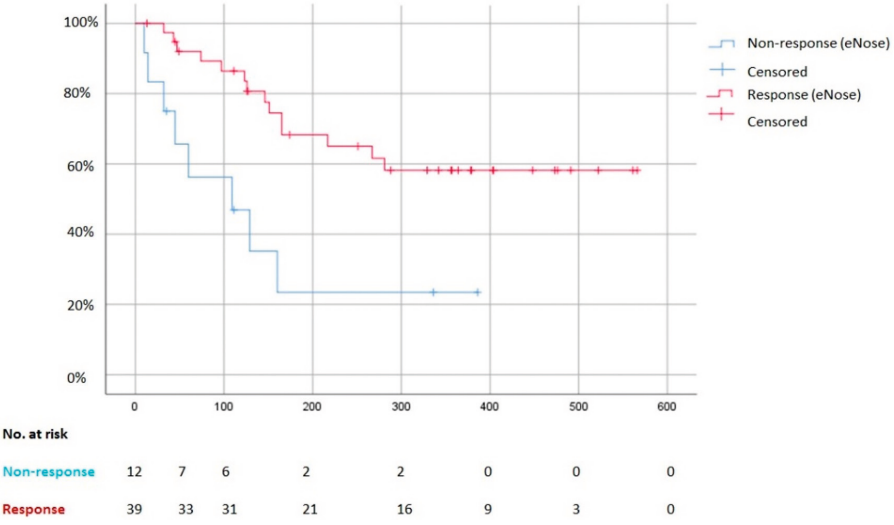
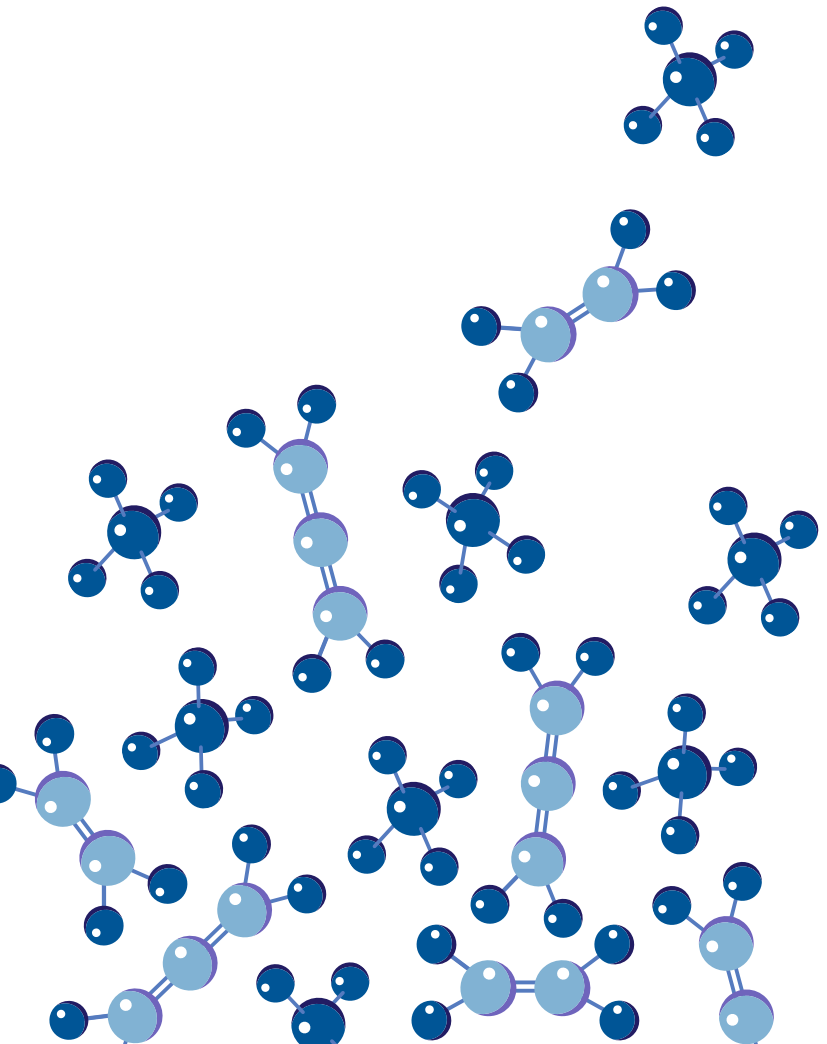


Figure S4B – Overall Survival: In this figure the overall survival (OS) analysis of the two groups based on the eNose test (with a cut-off of 0.72) of the validation set is shown with a Kaplan Meier curve. The median overall survival was 109 days (IQR: 32 - 160 days) for the non-responder group and not reached (IQR: 151 – NR) in the responder group, according to the eNose test. With a log rank (Mantel-Cox) test, the survival curve showed a significant difference between responders and non-responders according to their group based on the eNose (p=0.003).



6 *eNose analysis for early immunotherapy response monitoring in non-small cell lung cancer*

A. Buma*, M. Muller*, R de Vries, P.J. Sterk, V. van der Noort, M. Wolf-Lansdorf, N. Farzan, P. Baas, M. van den Heuvel

Lung Cancer. 2021;160:36-43.

*Contributed equally

Abstract

Objectives: Exhaled breath analysis by electronic nose (eNose) has shown to be a potential predictive biomarker before start of anti-PD-1 therapy in patients with non-small cell lung carcinoma (NSCLC). We hypothesized that the eNose could also be used as an early monitoring tool to identify responders more accurately at early stage of treatment when compared to baseline. In this proof-of-concept study we aimed to definitely discriminate responders from non-responders after six weeks of treatment.
Materials and Methods: This was a prospective observational study in patients with advanced NSCLC eligible for anti-PD-1 treatment. The efficacy of treatment was assessed by the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 at 3-month follow-up. We analyzed SpiroNose exhaled breath data of 94 patients (training cohort n=62, validation cohort n=32). Data analysis involved signal processing and statistics based on Independent SamplesT-tests and Linear Discriminant Analysis (LDA) followed by Receiver Operating Characteristic (ROC) analysis.
Results: In the training cohort, a specificity of 73% was obtained at a 100% sensitivity level to identify objective responders. The Area under the Curve (AUC) was 0.95 (CI: 0.89-1.00). In the validation cohort, these results were confirmed with an AUC of 0.97 (CI: 0.91-1.00).
Conclusion: Exhaled breath analysis by eNose early during treatment allows for a highly accurate, non-invasive and low-cost identification of advanced NSCLC patients who benefit from anti-PD-1 therapy.

Keywords: non-small cell lung cancer, immunotherapy, exhaled breath analysis, non-invasive biomarker.

Introduction

The recent introduction of immune checkpoint inhibitors (ICIs) in daily clinical practice has significantly improved the 5-year survival rate in patients with metastatic non-small cell lung cancer (NSCLC) [248]. Nevertheless, results have shown that only a minority of patients experiences a relevant clinical benefit [249]. Treatment continuation is currently based on tumor dynamics evaluated by radiological imaging. However, tumor dynamics can be difficult to interpret when tumor regression occurs slowly, there is no measurable disease, or tumors even transiently progress due to inflammation [250, 251]. Since the only validated predictive biomarker tumor PD-L1 expression is fairly inaccurate, other, and preferably non-invasive, predictive biomarkers are being investigated to avoid losing valuable time and undesirable immune-related adverse events (IRAEs), and to reduce unnecessary costs [249, 252-256].

Recent studies have been exploring the use of exhaled breath analysis with "electronic noses" (eNose), which recognize gas mixtures from volatile organic compounds (VOCs). VOCs are defined as chemical compounds that have a high vapor pressure at room temperature and are a result of metabolic changes in the body [229, 257]. ENoses have been designed for classification of VOCs by pattern recognition, which can be used for probabilistic assessment of disease states. Promising results have been observed in different diseases, particularly in the field of respiratory medicine [258, 259]. Recently, *De Vries et al.* showed that exhaled breath analysis by eNose can be used before start of treatment to identify NSCLC patients that show progressive disease (PD) to anti-PD-1 therapy with 100% specificity. This way, ineffective treatment could potentially be avoided in a quarter of the patients without withholding it to those who may benefit [260]. However, still a relevant proportion of patients will ultimately not benefit. An early monitoring tool for response during treatment would be helpful to identify those patients that are more likely to benefit from alternative options.

We hypothesized that exhaled breath patterns arising from metabolic/biochemical changes induced by effective anti-PD-1 therapy in patients with NSCLC can be used to discriminate true responders from non-responders more accurately at early stage of treatment when compared to baseline. According to this hypothesis, we expect that patients with a partial response (PR) will show greater metabolic/biochemical changes compared to patients with stable disease (SD) or PD, therefore resulting in a high predictive value in identifying true response to anti-PD-1 therapy when classifying patients with PR as responders and patients with SD or PD as non-responders. This proof-of-concept study therefore aims to determine the predictive value of exhaled breath analysis by eNose for the identification of advanced NSCLC patients with PR to anti-PD-1 therapy with 100% sensitivity after six weeks of treatment.

Methods

Study population

This was a prospective observational study in adult patients with advanced NSCLC eligible for treatment with anti-PD-1 therapy. Our cohort consists of two subsets of patients: 1) Patients included in the cohort of *De Vries et al.*, who also had received a second SpiroNose measurement after six weeks of treatment (n=64), and 2) patients recruited after publication who were only treated with pembrolizumab (n=30) [260]. Patients were recruited from the thoracic oncology outpatient clinic at the Netherlands Cancer Institute (NKI) in Amsterdam and the Radboudumc Hospital in Nijmegen between August 2015 and June 2019. The patients were only included if they had received SpiroNose measurements both at baseline (defined as 0-6 weeks prior to treatment start) as well as after six weeks of treatment (defined as 4-8 weeks of treatment), and received treatment in accordance with recent literature and local guidelines [183]. Details about the "full eligibility criteria for treatment with immunotherapy in NSCLC patients" have been described by *De Vries et al. [261]*. Patients received Nivolumab or Pembrolizumab treatment every two or three weeks, respectively (Figure 1). Patients were excluded from the study if they had consumed alcohol 12 hours before the measurement, or when they were not willing to participate. Additional restrictions in eating, drinking, smoking and medication were not requested in order to make exhaled breath measurements applicable in daily clinical practice [65].

The study was approved by the ethics review board of the NKI. Details are described in the *Online Supplement*. Patients participating in the Thoracic Oncology Biobank provided written informed consent according to the Thoracic Oncology Biobank study protocol. Measurements were not performed in patients with severe shortness of breath, inability to perform a vital capacity maneuver, or inability to hold breath for five seconds. Patients were only included if they had received SpiroNose measurements both at baseline (defined as 0-6 weeks prior to treatment start) as well as after six weeks of treatment (defined as 4-8 weeks of treatment). The choice for these cut-off periods was based on our aim to include as many patients as possible.

The ethics review board of the Netherlands Cancer Institute (NKI) concluded in writing that Dutch legislation on human participation in research was not considered to be applicable, given the non-invasive nature of this study that merely added exhaled breath analysis to standard diagnostic procedures. Other clinical data (e.g. CT scan, blood tests and lung function) used in this study were collected for routine clinical practice and were subsequently handled by complying with the Dutch Personal Data Protection Act (WBP). Despite the waiver that was provided by the ethics review board, the purpose of adding the eNose to routine diagnostics was explained to the patients who all gave their oral consent.

Measurements

Within two weeks before start of treatment, blood tests and spirometry were done for toxicity monitoring, and repeated every 2 and 6 weeks respectively. For response monitoring a computed tomography (CT) scan was done within 2 weeks before start of treatment, 6 and 12 weeks after start of treatment, and repeated every 3 months. Based on the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 criteria, tumor

dynamics evaluated in all patients with CT-imaging at baseline and at 3-month follow-up were classified as partial response (PR), stable disease (SD) or progressive disease (PD) [105]. Patients classified as PR were categorized as objective responders, while patients showing SD or PD were categorized as non-responders.

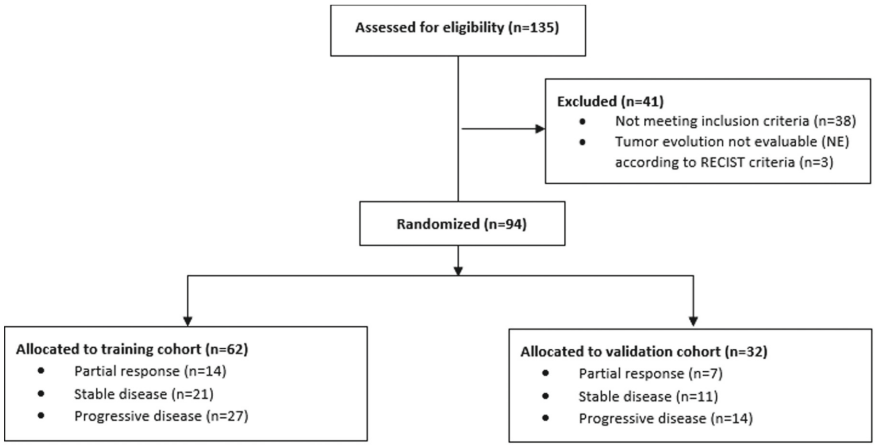


Figure 1 - CONSORT flow diagram of participants through the study.
The 38 participants that did not meet the inclusion criteria had not received SpiroNose measurements both at baseline (defined as 0-6 weeks prior to treatment start), as well as after six weeks of treatment (defined as 4-8 weeks of treatment).
Abbreviations: RECIST, Response Evaluation Criteria in Solid Tumors.

Study design

After response evaluation had been obtained for all patients, patients were randomized between a training and a validation cohort in a 2:1 ratio. Our aim was to keep both cohorts as representative as possible. Therefore, randomization was stratified according to the before mentioned response criteria at 3-month follow-up to keep an equal distribution in responses in both cohorts. Investigators were blinded to the exhaled breath data until after randomization.

In the training cohort two models for predicting response based on exhaled breath data were fitted: one using only baseline measurements (the “baseline model”) and one using both measurements collected at baseline and measurements collected after six weeks of treatment (the “on treatment model”). Then the performance of both models was evaluated in a cohort of patients not involved in the fitting of the models: the validation cohort.

Exhaled breath analysis

Exhaled breath measurements were performed using a cloud-connected eNose; SpiroNose® (Breathomix, Leiden, The Netherlands). The measurements took place the same day as the spirometry tests. The SpiroNose contains seven different cross-reactive metal oxide semiconductor (MOS) sensors and each sensor is present in duplicate both on the inside and outside of the SpiroNose. A detailed description of the SpiroNose measurement technology and breath sampling methods has been provided by *De Vries et al.* [234, 260]. The inner sensors measure the complete mixture of VOCs in exhaled breath and the outer sensors measure the ambient VOCs for background correction. Each sensor is used to determine two variables; 1) the highest sensor peak normalized to sensor 2, which is the most stable sensor, and 2) the ratio between the sensor peak and breath hold (BH) point [239, 261]. Measurements were performed in duplicate with a 2-minutes interval at baseline and after six weeks of treatment.

Data processing

Processing of the eNose sensor signals included filtering, de-trending, ambient correction and peak detection by the standard eNose software as described by *De Vries et al*. [65, 234, 260]. A .csv file was used to store the selected parameters (sensor peak- and peak/BH ratios) resulting from the signal processing and served as the source document for statistical analysis.

Statistical analysis

Patient and tumor characteristics

Data-analysis was performed using MatLab (Version 2019b) and IBM SPSS Statistics (Version 26) and is explained in the *Online Supplement*.

Patient and tumor characteristics were described and compared between responders and non-responders, for both cohorts separately, considering a p-value <0.05 as statistically significant. Continuous variables were reported as means (SD) or medians (IQR) for normally and non-normally distributed data, respectively. Categorical variables were reported as ratios. Intergroup comparisons were performed using One-way ANOVA tests, Kruskal Wallis tests or Chi-squared tests.

Sample size calculation

Due to logistic reasons the total number of patients in the cohort was fixed. However, a calculation for the training and validation cohort was possible. Our aim was to make the training cohort as large as possible, while still having sufficient patients in the validation cohort to draw meaningful conclusions. We decided on forehand that a model developed in the training cohort would be considered successful if the two-sided 95% DeLong confidence interval around the AUC as established in the validation cohort would be entirely above 0.70, thus clinically relevantly far removed from the null-value of 0.5. Furthermore, our aim was to develop a biomarker in the training cohort that would be as accurate as the biomarker of *De Vries et al*. (which had an AUC of 0.85), to show the added value of "on treatment" breath profiles [260]. With this in mind, we decided to randomize the patients in a 2:1 ratio (training cohort n=62, validation cohort n=32) if simulations would show that 32 patients in the validation cohort would still yield sufficient accuracy. In order to determine (prior to randomization) whether a validation cohort of 32 patients

would yield sufficient accuracy to declare a marker as accurate as the one developed by *De Vries et al*. "successful" according to the above criterion, we randomly drew 10.000 virtual validation cohorts of 32 patient each, and computed the AUCs with confidence intervals of the actual *De Vries et al.* biomarker in each of these cohorts. The mean lower bound of these confidence intervals was 0.766, indicating that a 2:1 randomization would indeed yield a sufficiently large validation cohort according to our pre-specified criterion.

Exhaled breath analysis: training cohort
Since the model of *De Vries et al.* aims to identify patients classified as PD with 100% specificity, our training cohort was used to make a new predictive model based on baseline measurements only (the "baseline model") to identify patients classified as PR with 100% sensitivity. Furthermore, a predictive model was composed that included measurements performed both at baseline and after six weeks of treatment (the "on treatment model") to be able to determine the additional value of measurements performed early during treatment. Independent Samples T-tests and Linear Discriminant Analysis (LDA) were used to identify sensor values with the highest contribution to the discrimination of patients classified as PR and patients classified as SD or PD. For the baseline model, only baseline sensor values were included in both analyses. For the on treatment model, baseline sensor values and sensor values obtained after six weeks of treatment were included. The suffixes _6, _absdif and _reldif are used to indicate the sensor variables "value after six weeks of treatment", "absolute difference" and "relative difference", respectively. Details regarding the construction of the models are provided in the *Online Supplement*. Receiver Operating Characteristic (ROC) curves were constructed for the composed predictive models, and associated Area Under the Curves (AUCs) and specificities when focusing on a 100% sensitivity to identify patients classified as objective responders were calculated.

In the training cohort, two predictive models were composed to identify at baseline and after six weeks of treatment patients with an objective response to anti-PD-1 therapy with 100% sensitivity, respectively. Both models were examined in the validation cohort to test external validity. Variables that were considered for inclusion in the models were 1) all normalized sensor peaks and peak/ breath hold (BH) ratios at baseline, 2) all normalized sensor peaks and peak/BH ratios after six weeks of treatment, and 3) all absolute and relative differences between baseline and six weeks of treatment in sensor peaks and peak/BH ratios. Absolute sensor differences were calculated by subtracting sensor values measured after six weeks of treatment from sensor values measured at baseline for each sensor. Relative sensor differences were calculated by dividing the calculated absolute sensor differences by the sensor values measured at baseline for each sensor. The characters _6, _absdif and _reldif are used to indicate the sensor variables "value after six weeks of treatment", "absolute difference" and "relative difference", respectively.

Firstly, the values of all predictors were compared univariately between patients with a partial response (PR) and patients with stable disease (SD) or progressive disease (PD) by means of an Independent Samples T-test. Secondly, various multivariate models created by linear discriminant analysis (LDA) were considered. We decided to include

absolute over relative sensor differences in het multivariate models, since relative sensor differences are based on ratios and therefore more prone to potential errors. Since for any sensor, any two of the variables "value at baseline", "value after six weeks of treatment" and "absolute difference" are sufficient to compute the remaining variable, we barred these variables from entering the models with more than two at a time. Variable selection for the final LDA models was based on the performance of the predictors in the univariate and multivariate models in combination with our aim of including a maximum of six predictors into each model to reduce the risk of overfitting.

Exhaled breath analysis: validation cohort
For each patient in the validation cohort the values of the discriminant functions given by the baseline model and the on treatment model (composed in the training cohort) were calculated. These values were used to construct ROC curves in the validation cohort. To test external validity, the AUCs of these ROC curves were compared to the AUCs of the ROC curves composed in the training cohort and to the fixed boundary of 0.7. The predictive accuracy of both models was established in the validation cohort based on AUC and on specificity when focusing on a 100% sensitivity to identify patients with an objective response.

Finally, the discriminant scores calculated from each model for each patient were converted into prediction scores to facilitate interpretation in daily clinical practice. A cut-off point was selected in the training cohort to identify objective responders with 100% sensitivity after six weeks of treatment. This cut-off point was translated into a prediction score and examined in the validation cohort by calculating the associated sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Survival analyses with Kaplan Meier curves were performed to assess the relation between our results and survival.

Results

In total, 94 advanced NSCLC patients were enrolled in this study (Figure 1), from who 64 were included in the study of *De Vries et al*. [260]. They were randomly assigned to the training (n=62) or validation cohort (n=32) (Table 1), according to the before mentioned criteria, resulting in 62 patients in the training cohort and 32 patients in the validation cohort. All baseline measurements were performed with a median of 2 weeks (range: 0-6 weeks) before treatment. The follow-up measurements were performed with a median of 6 weeks (range: 4-8 weeks) and 12 weeks (range: 10-14 weeks) for the eNose and CT-scan respectively. In the validation cohort, a significant difference was seen in choice of treatment between the three groups (p=0.01). Patients showing SD or PD after three months of treatment were more often treated with Nivolumab compared to Pembrolizumab. No significant differences were seen in any other baseline characteristic between the three groups.

Table 1 - Baseline characteristics of the training and validation cohort.

	Training (n=62)			Validation (n=32)		
	PR (n=14)	**SD (n=21)**	**PD (n=27)**	**PR (n=7)**	**SD (n=11)**	**PD (n=14)**
Patient						
Age in years, mean (SD)	64.7 (7.4)	66.7 (7.9)	65.63 (9.3)	64.3 (9.4)	65.2 (7.2)	59.4 (9.8)
Gender (males), N (%)	6 (42.9)	13 (61.9)	14 (51.9)	3 (42.9)	6 (54.5)	8 (57.1)
BMI, median (IQR)	22.8 (20.7-27.2)	25.3 (23.7-30.4)	27.0 (23.2-31.4)	24.9 (22.7-28.3)	26.6 (23.9-31.9)	25.0 (20.4-26.8)
FEV1%, median[a] (IQR)	2.0 (1.6-2.3)	1.6 (1.4-2.3)	1.9 (1.5-2.6)	1.8 (1.7-2.0)	1.9 (1.6-2.0)	1.9 (1.6-2.4)
WHO performance[b] (≥2), N (%)	1 (7.7)	0 (0.0)	2 (7.4)	0 (0.0)	2 (20.0)	1 (7.1)
Ethnicity (Caucasian)[b], N (%)	13 (100.0)	18 (94.7)	25 (92.6)	7 (100.0)	9 (90.0)	14 (100.0)
Smoking status, N (%)						
Never smoker	0 (0.0)	2 (9.5)	2 (7.4)	1 (14.3)	2 (18.2)	2 (14.3)
Current smoker	3 (21.4)	6 (28.6)	6 (22.2)	2 (28.6)	2 (18.2)	5 (35.7)
Ex-smoker	11 (78.6)	13 (61.9)	19 (70.4)	4 (57.1)	7 (63.6)	7 (50.0)
Pack-years[c], median (IQR)	25.0 (15.0-40.0)	31.0 (7.0-45.0)	30.0 (20.0-50.0)	41.0 (32.5-52.0)	27.0 (10.5-32.0)	29.0 (15.0-40.0)
Tumor characteristics						
Histologyb, N (%)						
AC	9 (69.2)		19 (70.4)	3 (42.9)	6 (54.5)	8 (57.1)
SCC	1 (7.7)	4 (22.2)	6 (22.2)	2 (28.6)	4 (36.4)	2 (14.3)
Other	3 (23.1)	1 (5.6)	2 (7.4)	2 (28.6)	1 (9.1)	4 (28.5)
Mutation statusb, N (%)						
KRAS positive	6 (46.2)	9 (47.4)	11 (50.0)	1 (20.0)	3 (33.3)	4 (33.3)
EGFR positive	0 (0.0)	0 (0.0)	1 (4.5)	2 (40.0)	0 (0.0)	1 (9.1)
BRAF positive	1 (10.0)	0 (0.0)	1 (4.5)	0 (0.0)	0 (0.0)	0 (0.0)

Table 1 - Baseline characteristics of the training and validation cohort. *(Continued)*

	Training (n=62)			Validation (n=32)		
	PR (n=14)	**SD (n=21)**	**PD (n=27)**	**PR (n=7)**	**SD (n=11)**	**PD (n=14)**
PD-L1 expressiond, N (%)						
Negative (0%)	1 (7.1)	9 (42.9)	6 (22.2)	0 (0.0)	3 (27.3)	6 (42.9)
Weak positive (<50%)	0 (0.0)	1 (4.8)	4 (14.8)	0 (0.0)	1 (9.1)	2 (14.3)
Strong positive (>50%)	8 (57.1)	8 (38.1)	4 (14.8)	4 (57.1)	2 (18.2)	0 (0.0)
Unknown	5 (35.7)	3 (14.3)	13 (48.1)	3 (42.9)	5 (45.5)	6 (42.9)
Cancer stage III, N (%)	1 (7.1)	3 (14.3)	4 (14.8)	1 (14.3)	3 (27.3)	0 (0.0)
Treatment						
Current[e], N (%)						
Nivolumab	9 (64.3)	13 (61.9)	19 (70.4)	3 (42.9)	9 (81.8)	13 (92.9)
Pembrolizumab	5 (35.7)	8 (38.1)	8 (29.6)	4 (57.1)	2 (18.2)	1 (7.1)
Line of treatment, N(%)						
1st line	5 (35.7)	6 (28.6)	3 (11.1)	2 (28.6)	3 (27.3)	1 (7.1)
≥ 2 line	9 (64.3)	15 (71.4)	24 (88.9)	5 (71.4)	8 (72.7)	13 (92.9)

Abbreviations: BMI, body mass index; FEV1, forced expiratory volume in 1 second; WHO, world health organization; COPD, chronic obstructive pulmonary disease; GOLD: Global Initiative for Chronic Obstructive Lung Disease; AC, adenocarcinoma; SCC, squamous cell carcinoma; EGFR, epidermal growth factor receptor.

[a] One patient (SD) missing FEV1% at baseline in both training and validation cohort.

[b] Not available for all patients at baseline in both training and validation cohort, percentage shown in percentage of known cases.

[c] Six patients (PR n=1, SD n=3, PD n=2) missing pack-years in training cohort.

[d] Significant difference between PR, SD and PD in both training (p=0.03) and validation cohort (p=0.002).

[e] Significant difference between PR, SD and PD in validation cohort (p=0.01).

Exhaled breath analysis: training cohort

Sensor 3 (p=0.001), sensor 5_BH (p<0.001), sensor 3_6 (p<0.001), sensor 4_6 (p=0.05), sensor 2_BH_6 (p=0.04), sensor 5_BH_6 (p=0.005), sensor 3_reldif (p<0.001) and sensor 3_absdif (p<0.001) significantly differed between patients classified as PR and patients classified as PD or SD, which is shown in *Supplementary Figure S1* and *Supplementary Figure S2* for sensor 3_absdif. Results obtained from LDA are described in the *Online Supplement*.

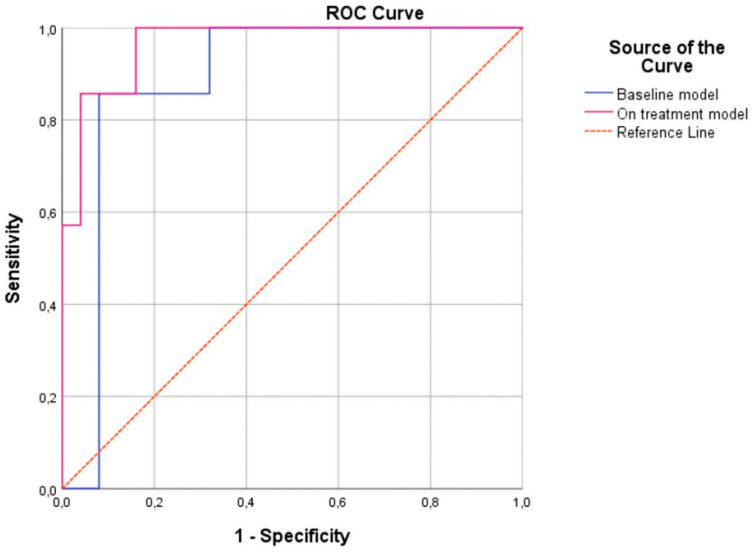


Figure 2 - ROC curves composed for the baseline model and model 2 predictive for the identification of patients showing an objective response to anti-PD-1 therapy in the validation cohort.
Baseline model: sensor 3, sensor 3_BH, sensor 5_BH and sensor 6_BH.
Model 2: sensor 4, sensor 6, sensor 1_6, sensor 6_6 and sensor 3_absdif.
Abbreviations: ROC, receiving operating characteristic; BH, breath hold; sensor 1_6, sensor value measured by sensor 1 after six weeks of treatment; sensor 6_6, sensor value measured by sensor 6 after six weeks of treatment; S3_absdif, sensor value difference between six weeks of treatment and baseline measured by sensor 3.

The first model (baseline model), based on baseline measurements only, reached a specificity of 54% when requiring 100% sensitivity and had a ROC-AUC of 0.81 (CI: 0.71-0.92) (Figure 2). In the second model (on treatment model), that included measurements performed both at baseline and after six weeks of treatment, a specificity of 73% at 100% sensitivity and a ROC-AUC of 0.95 (CI: 0.89-1.00) was reached (Figure 2). Details on the composition of the two models are provided in the *Online Supplement.*

Exhaled breath analysis: validation cohort

In the validation cohort, the baseline model reached a specificity of 68% when requiring 100% sensitivity and a ROC-AUC of 0.89 (CI: 0.76-1.00). The on treatment model reached a specificity of 84% at 100% sensitivity and a ROC-AUC of 0.97 (0.91-1.00). The baseline model reached a specificity of 68% at a ROC-AUC of 0.89 (CI: 0.76-1.00) (Figure 2 and *Supplementary Table S1*).

The equation below resulted from LDA on the on treatment model and can be used for response prediction in future patients. Details on the baseline model and the mathematical derivation of the models are provided in the *Online Supplement*.

Prediction score (patient) on treatment model =

$$\frac{1}{1 + e^{3.1180 + 4.6260 * S4 - 4.5276 * S6 + 1.0906 * S1\_6 - 0.9524 * S6\_6 + 26.0143 * S3\_absdif}}$$

The above equations were used to convert discriminant scores into prediction scores in the validation cohort (Figure 3). The cut-off point selected in the training cohort, aiming not to withhold anti-PD-1 therapy to objective responders after six weeks of treatment, corresponded to a prediction score of ≥0.14 for membership to the PR group. Patients with a prediction score below this cut-off point were classified as non-responders. In the validation cohort, this cut-off point showed a sensitivity of 100%, a specificity of 76%, a PPV of 54%, and a NPV of 100% to identify objective responders. Kaplan Meier survival curves are shown in *Supplementary Figure S3* and *Supplementary Figure S4*.

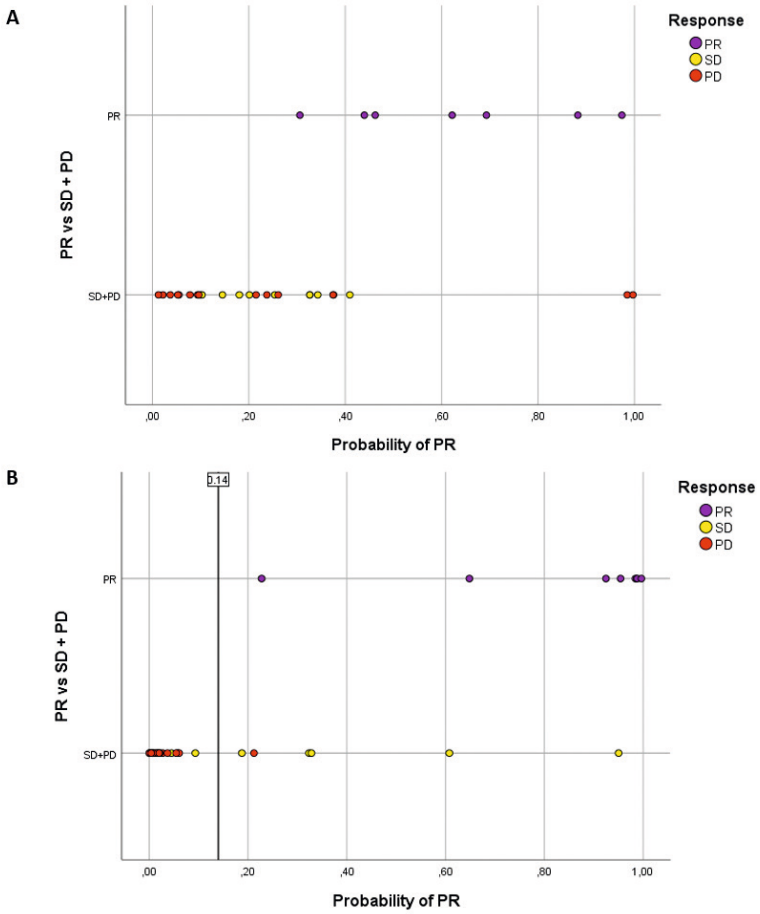


Figure 3 - Scatterplots representing the prediction scores calculated with the baseline model and model 2 for each patient (n=32) in the validation cohort.
A) Baseline model.
B) Model 2: All patients showing PR to anti-PD-1 therapy (n=7) are correctly classified when applying a cut-off point of ≥0.14 for membership to the objective response group. 5 out of 11 patients classified as SD and 1 out of 14 patients classified as PD are incorrectly classified.
Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease.

Discussion

This prospective observational study shows that SpiroNose exhaled breath analysis can be used to identify advanced NSCLC patients with an objective response to anti-PD-1 therapy more accurately at early stage of treatment when compared to baseline as part of routine assessment during early treatment monitoring in daily clinical practice. Results obtained in the training cohort were confirmed in the validation cohort with an AUC of 0.97 (CI: 0.91-1.00).

To the best of our knowledge, this is the first study that has applied eNose technology to identify advanced NSCLC patients with a PR to anti-PD-1 therapy and to investigate the potential additional value of SpiroNose exhaled breath measurements early during treatment. Our study extends the work of *De Vries et al.*, who investigated whether sensitivity to anti-PD-1 therapy in patients with advanced NSCLC might be reflected by a distinct exhaled breathprint. They showed that SpiroNose exhaled breath analysis could indeed be used to discriminate at baseline patients showing PD from patients showing PR or SD, and with a superior predictive performance than obtained with the current clinical standard biomarker PD-L1.

In this study, we obtained an increased discriminative potential for the identification of patients with PR when applying the on treatment model when compared to the baseline model (*Supplementary Table S1*). After six weeks of treatment, patients classified as PR showed a distinct clustering of prediction scores towards higher probabilities of an objective response, while patients classified as PD showed a distinct clustering towards lower probabilities of an objective response when compared to baseline. Patients categorized as SD, on the other hand, showed an increased spread in prediction scores, with the majority of scores falling back to low probabilities of an objective response (Figure 3). Based on these results, one could argue that this increased discriminative potential after treatment initiation might be partly driven by VOCs that arise from metabolic/biochemical changes induced by anti-PD-1 therapy. This would imply that a direct treatment effect could be monitored through exhaled breath. We therefore suggest that the on treatment model could therefore not only be used as a predictive biomarker to identify patients exhibiting primary resistance as early as six weeks following treatment initiation, but also as a real-time monitoring tool for therapeutic efficacy during follow-up to identify patients developing secondary resistance during course of treatment. However, we suggest this model first to be validated in an external, prospective, and preferably multicenter, validation cohort to confirm the predictive value obtained in our study. Subsequently, application of the model in daily clinical practice should be investigated in combination with other current biomarkers (e.g. clinical condition, serum markers, radiological imaging, histopathology, etc.) to help therapeutic decision-making during course of treatment. We expect that this approach will allow for an earlier and more precise identification of non-responding patients during anti-PD-1 therapy when compared to current follow-up care, and subsequently help to avoid undesirable events of treatment and losing valuable time in a higher percentage of these patients.

As eNoses have been designed for probabilistic assessment of VOCs, based on pattern recognition algorithms, It remains to be determined which specific metabolic/biochemical pathways contribute to the associations between the measured VOC-patterns and the patient response evaluation. When looking at our composed predictive models and the model of *De Vries et al.*, we observe that sensor 3, which is most sensitive to methane and natural gas, consistently has a major contribution to the discriminative performance of all models [260]. The model of *De Vries et al.* aimed to identify patients showing PD with a 100% specificity, classifying patients as PR or SD as responders. Our study aimed to improve the applicability of SpiroNose exhaled breath analysis in daily clinical practice. Mean sensor values calculated for sensor 3 for each response group showed that patients classified as PR had the highest mean sensor value at baseline, while patients with PD had the lowest (*Supplementary Table S3* and *Supplementary Table S4*). Furthermore, the SD group showed a mean sensor value more similar to the mean sensor value calculated for patients with PR. After six weeks of treatment, however, exhaled breath patterns distinctly differed for each response group. Patients categorized as PR showed a significant decrease in measured sensor values and had the lowest mean sensor value, while patients with PD had the highest. Interestingly, patients classified as SD showed a mean sensor value more similar to the mean sensor value calculated for the PD group. One could therefore speculate that the majority of patients classified as SD exhibit slow tumor progression during treatment. Classifying patients with PR or SD as responders during treatment would therefore have resulted in a smaller difference in mean sensor value between the responder and non-responder group, resulting in a lower predictive value in identifying true response to anti-PD-1 therapy when applying the on treatment model. We therefore believe that the current classification of responders brings the evaluation of response closest to the actual response occurring within the patient during treatment. Applying a predictive model that aims to identify VOC pattern changes occurring in true responders could therefore be more sensitive in identifying patients within the SD group who have a delayed but potentially durable response to anti-PD-1 therapy. This way, patients could be classified as responders or non-responders instead of PR/SD/PD, facilitating the decision to continue, stop or adapt treatment during current and future immunotherapy options [262]. Furthermore, we suggest that it should be investigated which individual VOCs contribute to the discrimination between responders and non-responders in order to draw conclusions about which specific metabolic/biochemical pathways are associated with response. Insight into these molecular mechanisms could then be used to improve the SpiroNose as biomarker tool.

One could argue that pattern recognition rather than identification of VOCs is an intrinsic limitation of using eNoses. Exhaled breath analysis technology comprises multiple methods for breath sampling [229]. Different studies have been able to identify multiple compounds associated with lung cancer by using methods that aim to detect, identify, and quantify specific, individual chemical compounds in exhaled breath [263-265]. However, these methods have shown some practical disadvantages, which makes them less suitable for clinical implementation yet [65, 228, 239, 260, 266]. In the present study, we were able to accurately identify true responders to anti-PD-1 therapy by using an eNose based on cross-reactive nonspecific sensor arrays without requiring

any restrictions except from alcohol consumption12 hours before the measurement. In addition, we were able to identify true responders in a relatively heterogenous population of patients (Table 1). This could imply that the SpiroNose identifies breath patterns associated with response that are not influenced by baseline characteristics and lifestyle of patients, which increases its external validity. Since it remains unclear which intrinsic and extrinsic factors determine the breath print typically for a response to immunotherapy and which set of VOC's characterize this breath print, we believe that the use of an eNose based on pattern recognition allows for a less error-prone, and therefore more accurate, approach for identifying responding patients as part of routine assessment in daily clinical practice when compared to individual VOC detection methods.

A limitation of our study might be the response categorization based on conventional radiological response criteria. Pseudoprogression, which is defined by a transient increase in tumor burden followed by a delayed decrease in tumor size, is considered one of the unusual response patterns when assessing efficacy of immunotherapy by radiological imaging and might result in incorrect classification of a subset of responders [251, 267]. Since the incidence of pseudoprogression in NSCLC patients is thought to be less than 5%, we believe the risk of misclassification bias in our study to be extremely low [267].

Conclusions

In conclusion, results obtained in the present study show that exhaled breath analysis by eNose allows for a highly accurate, non-invasive and low-cost identification of advanced NSCLC patients with an objective response to anti-PD-1 therapy as part of a routine assessment during early treatment monitoring in daily clinical practice. The clear advantage of such an identification is that application of ineffective treatments can be avoided in a higher percentage of non-responding patients, thereby preventing undesirable events and reducing unnecessary costs. Importantly, our study also paves the way for optimizing the clinical application of eNose exhaled breath analysis in patients with advanced NSCLC.

Supplemental material

The supplemental material:

Content included in this thesis:
Supplemental results
Figure S1 - Sensor value sensor 3 between baseline and six weeks of treatment
Figure S2 - Sensor value after six weeks of treatment
Figure S3 - Kaplan Meier curve showing the overall survival (OS)
Figure S4 - Kaplan Meier curve showing the progression free survival (PFS)
Table S1 - Predictive performance of the SpiroNose in both training and validation cohort.
Table S2 - Sensor values measured by sensor 3 and prediction scores
Table S3 - Mean sensor values calculated for S3, S3_6, and S3_absdif
Table S4 - Mean sensor values and difference in mean sensor values calculated for S3, S3_6, and S3_absdif

Results

Exhaled breath analysis: training cohort
Sensor 3, sensor 4, sensor 5_BH, sensor 6_BH, sensor 1_6, sensor 6_6, sensor, sensor 3_BH_6 and sensor 6_BH_6 resulted to be the sensors with the highest discriminative potential based on LDA when including both baseline sensor values and sensor values measured after six weeks of treatment. ROC curves constructed for sensor 3_reldif and sensor 3_absdif showed a large overlap between both curves, justifying our decision to include only absolute sensor signal differences when performing LDA. The best predictive model (model 1) was composed with sensor 4, sensor 6, sensor 3_BH_6, sensor 5_BH_6, sensor 6_BH_6 and S3_absdif and showed a specificity of 77% when focusing on a 100% sensitivity to identify patients classified with an objective response at an Area Under the Curve (AUC) of 0.97 (CI: 0.92-1.00). Model 2 consisted of sensor 4, sensor 6, sensor 1_6, sensor 6_6 and sensor 3_absdif. This model reached a specificity of 73% at an AUC of 0.95 (CI: 0.89-1.00) when focusing on a sensitivity of 100% to identify objective responders. Model 3 was composed with sensor 3_BH_6, sensor 5_BH_6 sensor 6_BH_6, and sensor 3_absdif, and reached a specificity of 71% at an AUC of 0.96 (CI: 0.91-1.00) (*Supplementary Table S1*).

When including only baseline sensor values, LDA showed a high contribution of sensor 3, sensor 3_BH, and sensor 6_BH to the discrimination between patients classified as PR and patients classified as SD or PD. Since the performed Independent Samples

T-test showed that sensor 3 and sensor 5_BH significantly differed between patients showing PR and patients showing SD or PD at baseline, our predictive baseline model was composed with sensor 3, sensor 3_BH, sensor 5_BH and sensor 6_BH. This model reached a specificity of 54% at a ROC-AUC of 0.81 (CI: 0.71-0.92) when focusing on a 100% sensitivity to identify patients classified as objective responders (*Supplementary Table S1*).
Exhaled breath analysis: validation cohort
In the validation cohort, prediction scores were used to visualize the discriminative potential for the baseline model and model 2 ("on treatment model"). In the context of constructing a ROC curve, the calculated prediction scores contain exactly the same information as the discriminant scores [260].
However, prediction scores facilitate interpretation in daily clinical practice.

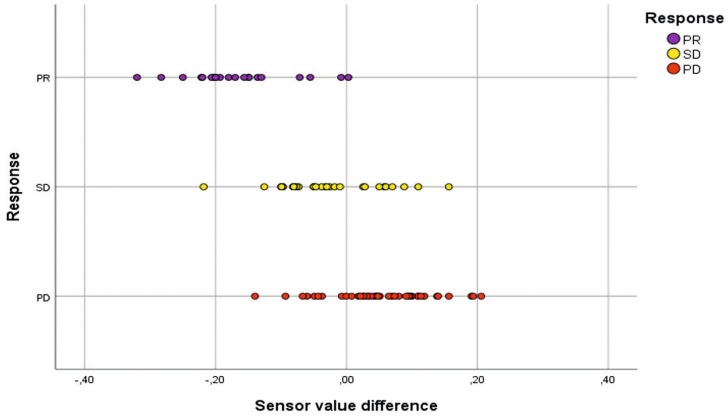


Figure S1 - Sensor value differences measured by sensor 3 between baseline and six weeks of treatment for each patient (n=94). 20 out of 21 patients classified as PR, 21 out of 32 patients classified as SD, and 8 out of 41 patients classified as PD show decreased sensor values after six weeks of treatment when compared to baseline. One patient classified as PR and one patient classified as PD show stable measured sensor values over time. All other patients show an increase in measured sensor value. Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease.

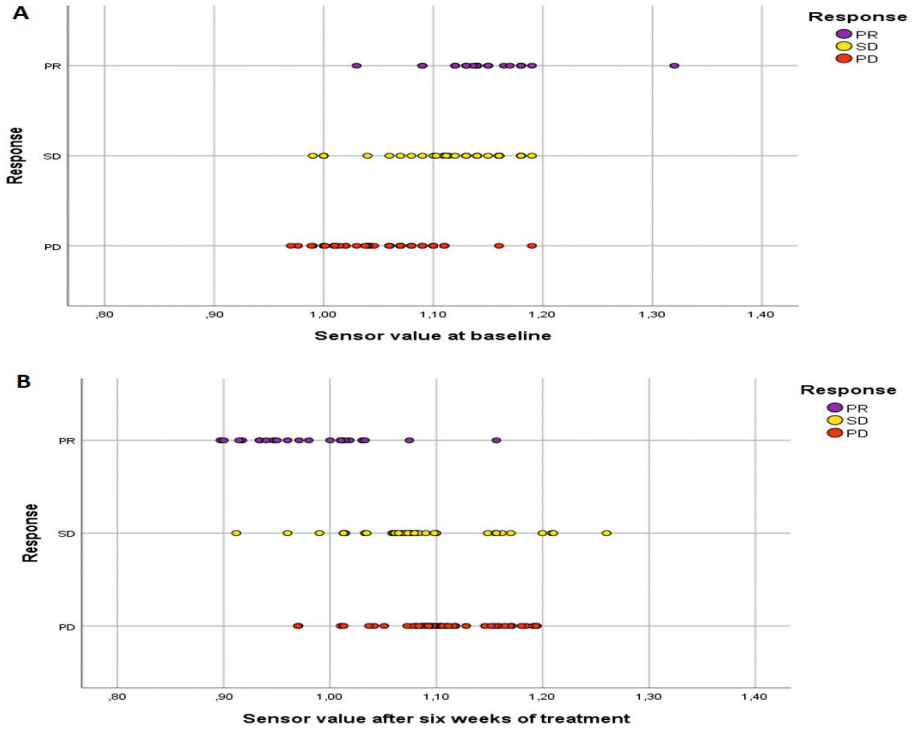
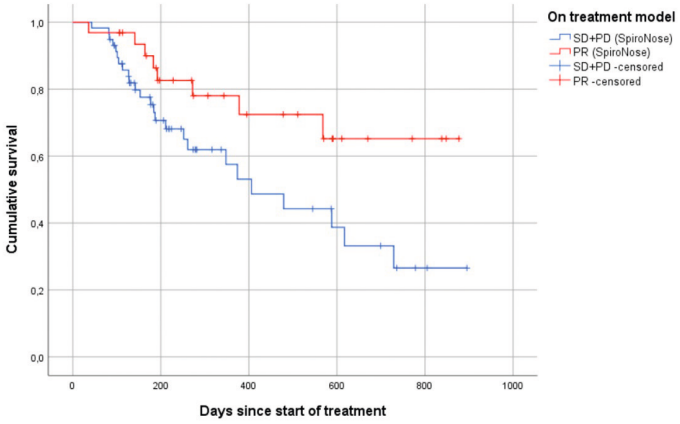


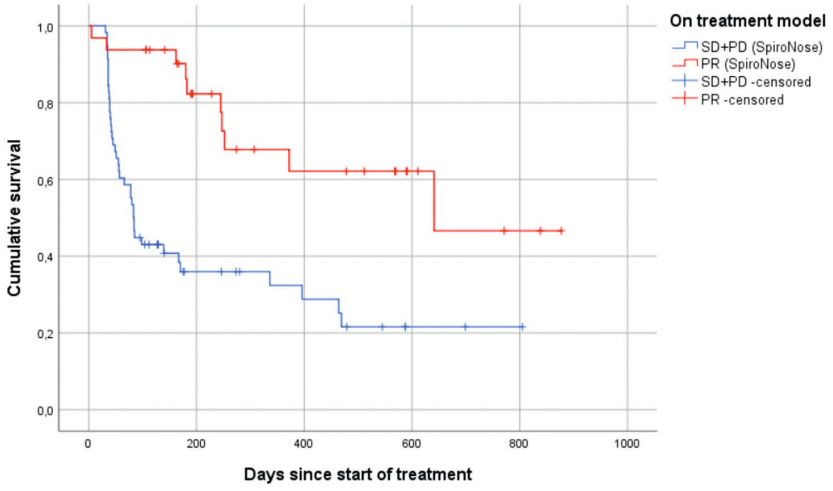
Figure S2 - Sensor values measured by sensor 3 at (A) baseline and (B) after six weeks of treatment for each patient (n=94). Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease.



No. at risk	0	200	400	600	800	1000
PR	32	20	12	6	3	0
SD+PD	58	29	12	7	2	0

Figure S3 - Kaplan Meier curve showing the overall survival (OS) analysis of the two groups (PR vs. SD+PD) when applying the on treatment model (with a cut-off of 0.14) of 90 advanced NSCLC patients.

The mean overall survival was 672 days (CI: 553-791 days) for patients classified as PR and 482 days (CI: 378-585 days) for patients classified as SD or PD, according to the SpiroNose test. The survival curve showed a significant difference with a log rank (Mantel-Cox) test between objective responders and patients classified as SD or PD according to their group based on the SpiroNose (p=0.03). Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease; NSCLC, non-small cell lung cancer; CI, confidence interval.



No. at risk	0	200	400	600	800	1000
PR	32	18	11	5	2	0
SD+PD	58	13	8	2	1	0

Figure S4 - Kaplan Meier curve showing the progression free survival (PFS) analysis of the two groups (PR vs. SD+PD) when applying the on treatment model (with a cut-off of 0.14) of 90 advanced NSCLC patients.

The mean progression free survival was 586 days (CI: 449-724 days) for patients classified as PR and 275 days (CI: 189-362 days) for patients classified as SD or PD, according to the SpiroNose test. The survival curve showed a significant difference with a log rank (Mantel-Cox) test between objective responders and patients classified as SD or PD according to their group based on the SpiroNose (p<0.001).

Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease; NSCLC, non-small cell lung cancer; CI, confidence interval.

Table S1 - Predictive performance of the SpiroNose for the identification of patients showing an objective response to anti-PD-1 therapy in both training and validation cohort.

Model	N	Cut-off point	Sensitivity (%)	Specificity (%)	AUC (95% CI)
Training					
Baseline model	14 vs. 48	0.1080	100	54	0.81 (0.71-0.92)
On treatment model	14 vs. 48	-0.2770	100	73	0.95 (0.89-1.00)
Validation					
Baseline model	7 vs. 25	-0.5794	100	68	0.89 (0.76-1.00)
On treatment model	7 vs. 25	-0.5575	100	84	0.97 (0.91-1.00)

Abbreviations: N, number; AUC, area under the curve; CI, confidence interval.

Table S2 - Sensor values measured by sensor 3 and prediction scores calculated with the baseline model and on treatment model for each patient (PR n=7, SD n=11, PD n = 14) in the validation cohort.

Patient ID	Response	S3	S3_6	S3_ absdif	P score baseline model	P score on treatment model	Difference[a]	OS[b]	PFS[b]
1	PR	1.18	0.90	-0.28	0.68	0.98	+0.30	478	478
2	PR	1.15	0.90	-0.25	0.88	0.99	+0.11	106	106
3	PR	1.14	1.01	-0.13	0.46	0.64	+0.18	189	189
4	PR	1.32	1.00	-0.32	0.97	1.00	+0.03	838	838
5	PR	1.14	0.91	-0.23	0.61	0.92	+0.31	771	771
6	PR	1.17	0.95	-0.22	0.42	0.95	+0.53	228	228
7	PR	1.13	1.07	-0.06	0.30	0.22	-0.08	591	591
8	SD	1.13	0.91	-0.22	0.36	0.95	+0.59	395	372
9	SD	1.00	1.16	+0.16	0.11	0.00	-0.11	895	469
10	SD	1.11	1.01	-0.10	0.25	0.32	+0.07	511	511
11	SD	1.08	1.07	-0.01	0.18	0.09	-0.09	588	588
12	SD	1.10	1.06	-0.04	0.14	0.18	+0.04	197	180
13	SD	1.12	1.07	-0.05	0.33	0.04	-0.29	479	479
14	SD	1.16	1.03	-0.13	0.31	0.60	+0.29	343	245
15	SD	1.16	1.08	-0.08	0.38	0.32	-0.06	183	182
16	SD	1.16	1.21	+0.05	0.31	0.02	-0.29	219	139
17	SD	1.07	1.10	+0.03	0.20	0.02	-0.18	778	396
18	SD	0.99	0.96	-0.03	0.11	0.06	-0.05	-	-
19	PD	1.04	1.10	+0.06	0.09	0.01	-0.08	406	36
20	PD	1.07	1.08	+0.01	0.26	0.03	-0.23	374	167
21	PD	1.00	1.19	+0.19	0.04	0.00	-0.04	112	34
22	PD	1.16	1.09	-0.07	0.36	0.21	-0.15	36	33
23	PD	1.11	1.07	-0.04	0.23	0.06	-0.17	132	39
24	PD	1.08	1.13	+0.05	0.99	0.01	-0.98	282	55
25	PD	1.08	1.11	+0.03	1.00	0.02	-0.98	211	57
26	PD	1.06	1.16	+0.10	0.05	0.01	-0.04	112	112
27	PD	1.08	1.04	-0.04	0.02	0.04	0.02	252	31
28	PD	1.09	1.12	+0.03	0.07	0.05	-0.02	127	36
29	PD	1.01	1.08	+0.07	0.01	0.00	-0.01	113	39
30	PD	1.00	1.09	+0.09	0.10	0.01	-0.09	206	43
31	PD	1.09	1.11	+0.02	0.21	0.02	-0.19	729	35
32	PD	1.04	1.15	+0.11	0.05	0.00	-0.05	187	51

Abbreviations: PR, partial response; SD, stable disease; PD, progressive disease; S3, sensor value measured by sensor 3 at baseline; S3_6, sensor value measured by sensor 3 after six weeks of treatment; S3_absdif, absolute sensor value difference between sensor values measured by sensor 3 at baseline and after six weeks of treatment; P, prediction; OS, overall survival; PFS, progression free survival.

[a] *Difference between prediction scores calculated with the baseline model and on treatment model.*

[b] *Time unit is days.*

Table S3 - Mean sensor values calculated for S3, S3_6, and S3_absdif for each response group.

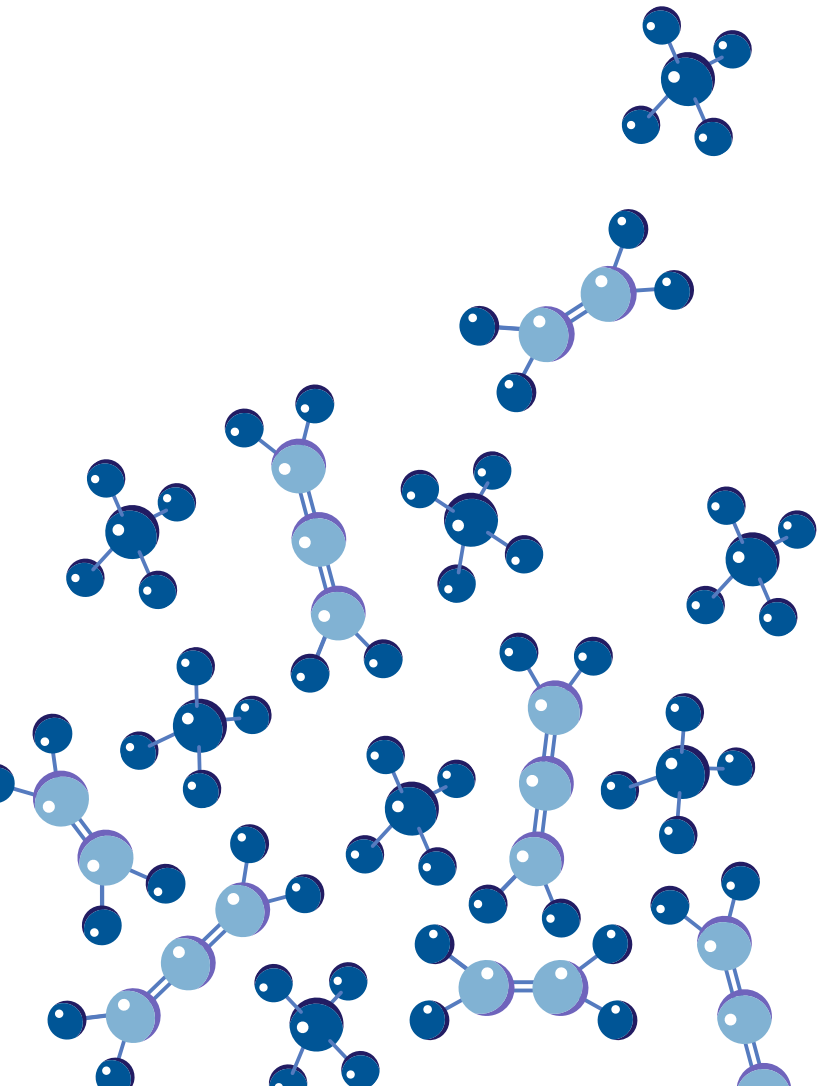
Advanced NSCLC (N=94)	
PR	(n=21_
S3	1.15
S3_6	0.98
S3_Absdiff	-0.17
SD	(n=32)
S3	1.11
S3_6	1.09
S3_Absdiff	-0.03
PD	(n=41)
S3	1.05
S3_6	1.10
S3_Absdiff	+0.05

Abbreviations: S3, sensor value measured by sensor 3 at baseline; S3_6, sensor value measured by sensor 3 after six weeks of treatment; S3_absdif, absolute sensor value difference between sensor values measured by sensor 3 at baseline and after six weeks of treatment; NSCLC, non-small cell lung cancer; PR, partial response; SD, stable disease; PD, progressive disease.

Table S4 - Mean sensor values and difference in mean sensor values calculated for S3, S3_6, and S3_absdif for the group of objective responders and the group of patients classified as SD or PD.

	Advanced NSCLC (n=94)		
	PR (n=21)	**SD+PD (n=73)**	**Difference**
S3	1.15	1.08	+0.07
S3_6	0.98	1.10	-0.12
S3_absdiff	-0.17	0.02	-0.18

Abbreviations: S3, sensor value measured by sensor 3 at baseline; S3_6, sensor value measured by sensor 3 after six weeks of treatment; S3_absdif, absolute sensor value difference between sensor values measured by sensor 3 at baseline and after six weeks of treatment; NSCLC, non-small cell lung cancer; PR, partial response; SD, stable disease; PD, progressive disease.



7 I love my dog, editorial

M. Muller, P. Baas.

Lung Cancer 2019. 135; 228-229

In many households, a dog is part of the family. The dog protects you, makes sure you will get your exercise while taking him for a walk and if lucky, and gets the paper for you on a Saturday morning. Nowadays dogs are more than just a pet, mostly because of their well-known good scent. By entering the US or Australia, dogs are used to identify drug-carriers or persons trying to import food.

People use a dog's outstanding sense of smell in many ways. The olfactory capability of dogs is about 10.000-100.000 times better than human beings. Whereas the human only has 5 million scent receptors, the nose of a bloodhound contains about 300 million olfactory receptors[268]. Even with a low limit of detectability for volatile organic compounds (VOCs) of one part per trillion (1×10^{12}), dogs are able to smell multiple different odors [269, 270]. Dogs are already being used in daily clinical practice. In the United States, Dogs4Diabetics train dogs for early recognition of hypoglycemia, so that a patient can take action in time. Some dogs are trained to recognize the start of an epileptic insult, to help a patient finding a better and safer environment.

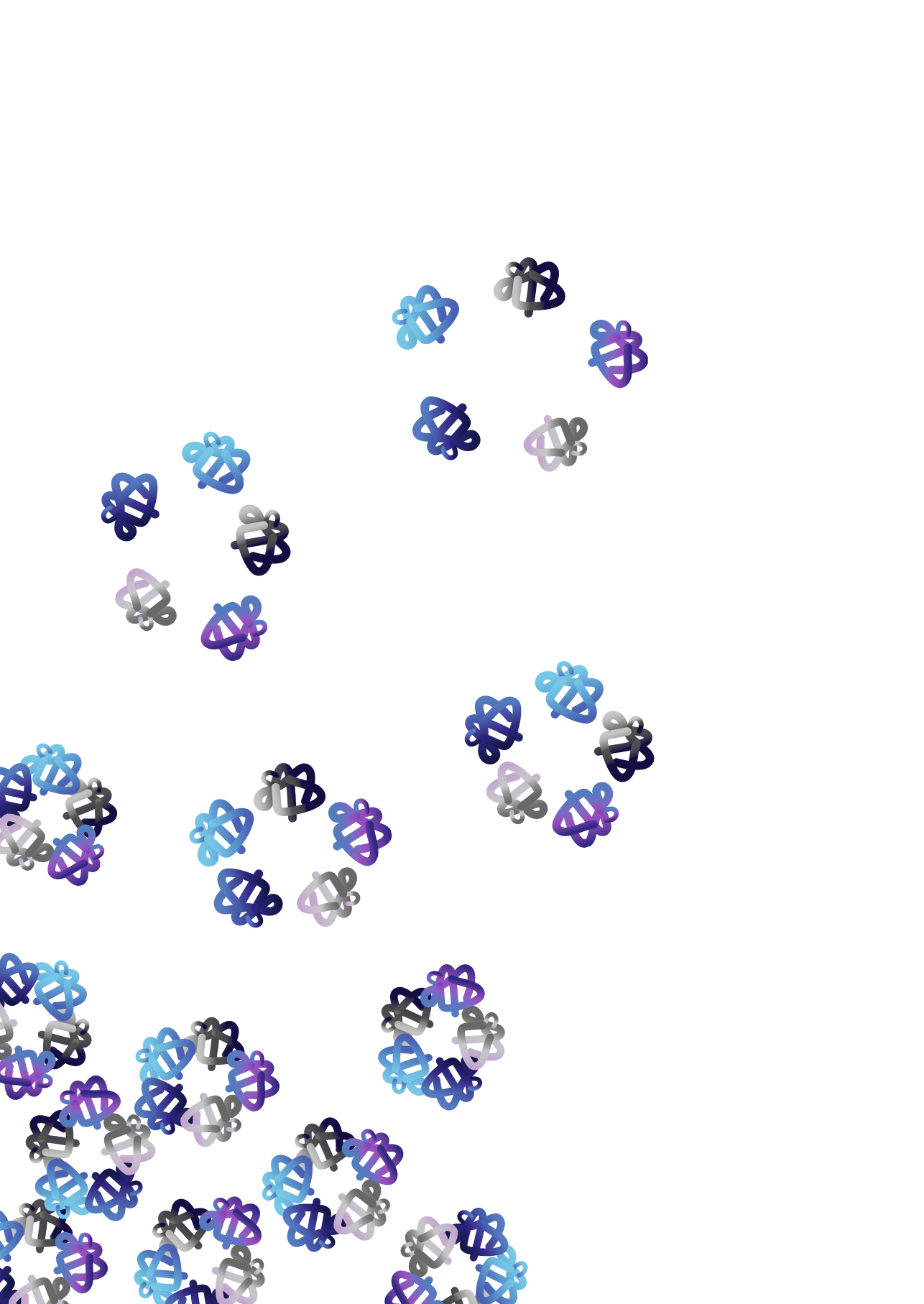
In the last decades, research is committed to basically copy the nose of a dog by developing electronic devices, so called electronic noses (e-nose), for the detection and recognition of odors and flavors. Exhaled breath of humans contains thousands of VOCs that originate from metabolic processes[227] , which can be detected with such devices. Multiple studies show that with the analysis of exhaled breath it is possible to differentiate between patients with different diseases like COPD, asthma, lung cancer compared to healthy subjects [65, 271-273]. Still, it has its limitations. One of the theories how the human scent works, is that we are not able to detect individual VOCs per se, but recognize a pattern. One example is that we recognize an orange as an orange and can differentiate it from a clementine, although we do not know which VOCs cause that specifically. However, in general the difficulty is that with all the current methods, we are not able to find and validate a pattern in multiple studies. To analyze VOCs by molecule GC-Mass Spectroscopy is required. This is complicated, expensive, time consuming and needs data analysis. Also, in multiple studies, multiple different important VOCs are found [274]. An alternative is to analyze multiple gasses simultaneously by using special sensors. This approach has gained popularity, but due to lack of standardization, studies cannot easily be compared.

In a previous paper, Guirao et al [275] presented the use of a dog trained identify patients with lung cancer. He was able to identify patients with lung cancer smelling their exhaled breath. Besides lung cancer, there were two control groups: patients with lung diseases like COPD and healthy subject. The results were impressive with a sensitivity of 95% and a specificity of 98%. Using exhaled breath analysis would be of great importance in diagnosing (early) disease and response monitoring. Until now there are no general accepted screening methods available in Europe. In the US screening programs are available, but the number needed to screen with CT scan is still 1:320 [276]. In the Netherlands, the NELSON-trial investigates the benefits of screening with a CT-scan compared to a control group. They found a survival benefit in the screened group, which was 26% better after 10 years for the male participants [277]. For females, the survival benefit was even better (41.8%), although not significant [278]. However, with a CT-scan, nodules are found where the distinction between benign and malignant is very difficult. The key question is if a dog can discriminate if a solitary pulmonary nodule is benign or malignant. Therefore, in their follow-up study published in the current issue

of Lung Cancer, they used the same dog for the discrimination between malignant and benign lesions [279]. Thirty patients participated in this study. Again, they reached a very high sensitivity and specificity, suggesting that the smell of lung cancer is quite specific.

Despite the benefits of the use of a dog, the question remains if this is a realistic option for clinical practice. For a start, you will need a dog from a breed which is known for its good scent. You will need to train and nurture it, which is a time consuming activity. And how do we deal with all the quality assurance? Is there a CE label for these dogs and how do we keep the quality high? Dogs are living beings and are subject to infections or other diseases. Will the analytic quality changes during the day or by aging and how many working hours is a dog allowed to function? These and many more practical factors urge us to seek other avenues of which the eNose is very popular. Processes of standardization and real time availability of the outcome will lead to a rise in the use of these machines.

In conclusion, with their outstanding good scent dogs are able to smell the small differences between multiple odors. Research is ongoing to contrive the added value in clinical practice for the recognition of lung cancer. And for your dog at home: Yes, your dog is able to smell the difference between, for example, whether or not you took a shower, and despite that, he will remain faithful to you and still like you. Therefore, the next unfortunate time he eats your paper on a Saturday morning, please keep this in mind and think “Yeah, I love my dog (too)”.



8 Validation of a clinical blood-based decision aid to guide immunotherapy treatment in patients with non-small cell lung cancer

M. Muller, R. Hoogendoorn, R. Moritz, V. van der Noort, M. Lanfermeijer, Catharina M. Korse, D. van den Broek, J. J. den Hoeve, P. Baas, H.H. van Rossum, M. M. van den Heuvel

Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine. 2021;43(1):115-27.

Abstract

Background: The widespread introduction of immunotherapy in patients with advanced non-small cell lung cancer (NSCLC) has led to durable responses but still many patients fail and are treated beyond progression.
Objective: This study investigated whether readily available blood-based tumor biomarkers allow accurate detection of early non-responsiveness, allowing a timely switch of therapy and cost reduction.
Methods: In a prospective, observational study in patients with NSCLC treated with nivolumab or pembrolizumab, five serum tumor markers were measured at baseline and every other week. Six months disease control as determined by RECIST was used as a measure of clinical response. Patients with a disease control < 6 months were deemed non-responsive. For every separate tumor marker a criterion for predicting of non-response was developed. Each marker test was defined as positive (predictive of non-response) if the value of that tumor marker increased at least 50% from the value at baseline and above a marker dependent minimum value to be determined. Also, tests based on combination of multiple markers were designed. Specificity and sensitivity for predicting non-response was calculated and results were validated in an independent cohort. The target specificity of the test for detecting non-response was set at > 95%, in order to allow its safe use for treatment decisions.
Results: A total of 376 patients (training cohort: 180, validation cohort: 196) were included in our analysis. Results for the specificity of the single marker tests in the validation set were CEA: 98·3% (95%CI: 90·9–100%), NSE: 96·5% (95%CI: 87·9–99·6%), SCC: 96·5% (95%CI: 88·1–99·6%), Cyfra21·1 : 91.8% (95%CI: 81·9–97·3%), and CA125 : 86·0% (95%CI: 74·2–93·7%). A test based on the combination of Cyfra21.1, CEA and NSE accurately predicted non-response in 32.3% (95% CI 22.6–43.1%) of patients 6 weeks after start of immunotherapy. Survival analysis showed a significant difference between predicted responders (Median PFS: 237 days (95%CI 184–289 days)) and non-responders (Median PFS: 58 days (95%CI 46–70 days)) ($p < 0.001$).
Conclusions: Serum tumor marker based tests can be used for accurate detection of non-response in NSCLC, thereby allowing early and safe discontinuation of immunotherapy in a significant subset of patients.

Keywords: Serum tumor marker, CEA, Cyfra, SCC, NSE, CA125, Nivolumab, response, Longitudinal

1. Introduction

Immune checkpoint based therapies for lung cancer have changed the therapeutic landscape of and survival from non-small cell lung cancer (NSCLC) [3, 5, 14]. Unfortunately, still a limited number of patients respond to immune checkpoint based treatment and non-responsiveness remains a clinical challenge [3, 24]. Therefore, treatment monitoring in order to detect non-responsiveness is of key importance and rapid detection of non-responsiveness potentially allows a prompt next in line treatment initiation avoiding unnecessary side effects and costs.

For NSCLC follow-up several circulating tumor biomarker are available [280-282]. Most biomarkers available in clinical practice have not been validated as monitoring tools [281, 283] . Tumor markers readily available at medical laboratories and potentially useful to monitor NSCLC treatment response include CA125, carcinoembryonic antigen (CEA), cytokeratin 19 fragments (Cyfra 21·1), neuron-specific enolase (NSE), and squamous cell carcinoma antigen (SCC) [284-286]. Though evidence supports the clinical application of some of these tumor biomarkers for lung cancer, no clear guidance is available [280, 282, 284, 287, 288]. The interpretation of these tumor biomarkers, when used for monitoring a specific cancer treatment, is therefore generally based on expert opinion and personal experience.

Recently, a method and software package, called ReMarker, was developed to assess the applicability of tumor marker changes after start of treatment in the response assessment [289]. We used this application to design and validate biomarker-response based tests that allow an accurate and early detection of non-responsiveness to immunotherapy for patients with NSCLC. This would allow early discontinuation of ineffective therapy and provide a window of opportunity for initation of subsequent other treatment opportunities. Besides, it would reduce potential side effects and costs. Our aim was to define and clinically validate an early response tool that accurately predicts non-response and can be easily applied in daily clinical practice based on changes in tumor markers during therapy.

2. Methods

2.1. Study population

In a prospective, observational study, patients with NSCLC treated with nivolumab or pembrolizumab were included. Serum tumor markers CA125, CEA, Cyfra 21·1, NSE, and SCC were measured. Using the data on clinical outcome a test for every separate tumor marker was designed: our aim was to design and optimize a test that identifies non-responders (as determined by RECIST at six months) as early as six weeks after starting treatment based on their serum marker values. Each such test was defined as positive if the value of that tumor marker met two criteria: (i) elevation of 50% compared to baseline and (ii) above a minimum value. A training cohort was used to determine the optimal minimum value for each tumor marker. An independent validation cohort was used to validate the resulting tumor marker tests. Also the performance of combining the results of the individual tumor marker tests were evaluated. In this single-center study all patients with NSCLC who started their treatment between March 2013 and September 2018 in The Netherlands Cancer Institute were included. Follow-up was available until

January 2019. All consecutive patients receiving immunotherapy in a variety of settings, such as routine care, early access, compassionate use program, and clinical trials, were treated according to corresponding protocols. Patient criteria for receiving nivolumab treatment were previously described [24] and can also be found in the supplemental material, as are the pembrolizumab criteria. If a patient had received immunotherapy in two different treatment lines, the initial treatment line was taken. Tumor markers were measured at baseline and prior to each consecutive cycle together with other routine blood assessment tests as standard of care. The monitoring of response was done with a CT scan before start of treatment, and after 6 weeks, 3 months and every 3 months thereafter. Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 were used, accordingly progressive disease (PD), stable disease (SD), and partial response (PR) [105]. Patients who were progressive before the endpoint of six months, were classified as having no clinical benefit (NCB), as previous described in Rizvi et al. [112]). Our study was approved by the local medical ethical committee (PTC NKI-AvL, NL45524.031.13), patient privacy committee and performed according to the institutional patient privacy protocols. In January 2017 all patients who had been treated with nivolumab at that time were randomly assigned to the training or validation cohort in a 2:1 ratio (Figure 1), as described in the sample size calculation (supplemental material). The training cohort was used to make and refine the ReMarker application (see below). After this randomization, no more patients were added to or removed from this training cohort. The validation cohort consisted of patients who were initially randomized to the validation cohort, and those who started their nivolumab treatment after January 2017 or who were treated with pembrolizumab, up to a 1:1 ratio.

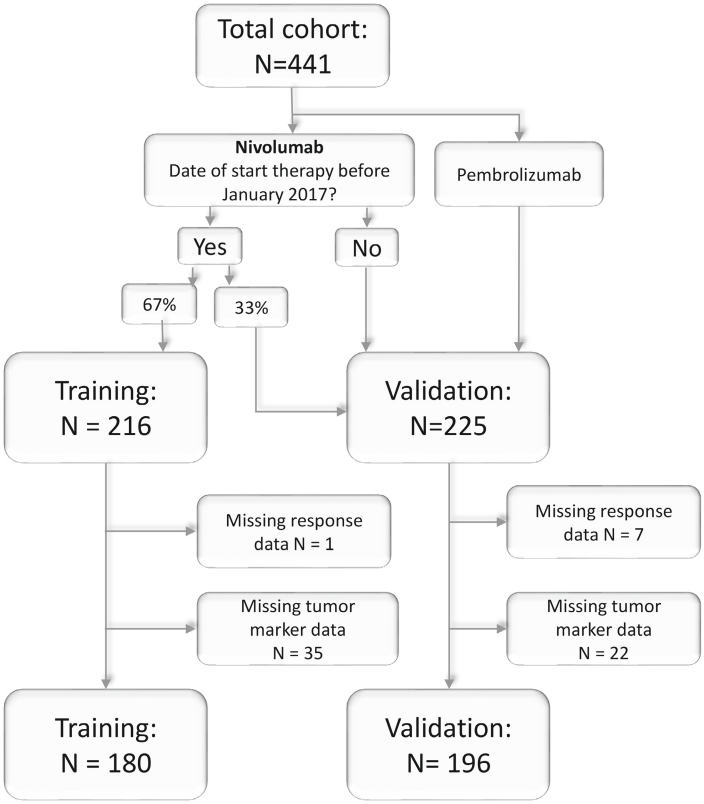


Figure 1 - Consort chart.
All patients treated with immunotherapy as second or higher line in the training and validation cohort.

2.2. Design of tumor marker test

Analysis of the obtained serum samples were performed on a daily (CA125,CEA, Cyfra 21·1, and NSE) or twice weekly (SCC) basis. CA125, CEA, Cyfra 21·1, and NSE were measured using a Cobas 6000 system (Roche diagnostics) and SCC was measured on a Kryptor system (Thermo Fisher), both according to the manufacturer's instructions. The applied reference ranges for the tumor markers were < 20 U/ml for CA125 (< 35 U/ml for premenopausal females), < 6g/L for CEA, < 1·9g/L for Cyfra 21·1, < 12·5g/L for NSE and < 2·0g/L and < 1·5g/L for SCC for males and females, respectively. The application ReMarker was used to study multiple time points and multiple cut-offs. The correlation to clinical response was visualized in Biomarker Response Characteristic plots (BReC plots) (Fig. S1) [289]. The baseline measurement was defined as minus 3 weeks until 0 weeks before start of treatment. A follow-up time point of 6 weeks was designated as primary optimization follow-up time, since in our practice this is the first clinical evaluation moment for response evaluation. This follow-up time point was defined as a measurement 5 or 6 weeks (±3 days) after start of treatment. If there was more than one

measurement in one of these periods, the latest measurement was taken. The training set was used to optimize the test per single tumor marker for the prediction of non-response, which was defined as PD, NCB or deceased after six months of immunotherapy treatment. The other patients were classified as responders. The following factors were taken into consideration for the design (and are also explained in Table S1): (I + II) In order to obtain an easy-to-calculate test, we defined our test as positive (i.e. predictive for non-response) when the marker increased with 50% from baseline and was above the marker dependent minimum value (Fig. S1); (III) The minimum value criterion was applied to exclude patients with small biomarker increases at low concentrations that results in large relative increases thereby reducing the effect of (pre-) analytical and biological "noise"; (IV) The optimal minimum value per marker was determined by calculating the specificity and sensitivity (Fig. S2); (V) Minimum values yielding a specificity of ≥ 97·5% in the training set per individual markers were considered a good cut-off; (VI) Minimum values yielding a sensitivity of > 20% were considered a good cut-off. An overview of the considerations can be found in Table S1. For each tumor marker we chose a minimum value satisfying the criteria in the training cohort (Fig. S2). Then, in the validation cohort, the sensitivity, specificity, positive predicted value (PPV) and negative predicted value (NPV), all with a 95% confidence interval, were calculated for the resulting test per tumor marker. After the best test per single tumor marker had been determined, the combination of tumor markers was tested in the training cohort (Table S2). A test was considered positive if at least one of the tumor markers increased with 50% above baseline. Only the tests in the training cohort that fulfilled abovementioned criteria were validated in the validation cohort, again in terms of sensitivity, specificity, NPV, and PPV, all with a 95% confidence interval. The performance of the tests was also investigated from week 2 until week 20, with biweekly tests, for both the training cohort and validation cohort in order to allow a more general application. Survival analyses and cox-regression analyses were performed assessing the predictive value of the tests for overall survival (OS) and progression free survival (PFS). OS was defined as the number of days between the day of start of treatment and date of death, PFS as the number days between the day of start of treatment and date of progression or death, whichever came first. SPSS (v25; SPSS, Chicago, USA) was used for the descriptive statistics. Descriptive statistics were expressed as mean ± SD if data were normally distributed and as median (interquartile range) if data were non-normally distributed. Between group comparisons were performed using Mann-Whitney U tests, two sample unpaired t-tests of Chi-Squared tests. From all the patients with a false-positive result, the medical record was checked for possible confounders.

Furthermore, a small cohort with patients who were treated with pembrolizumab in first line (rather than second line) was also available for analysis.

3. Results

3.1. Patients

A total of 441 patients were included in our study, 216 in the training set and 225 in the validation set (Table 1). From these patients, 389 patients were treated with nivolumab and 52 with pembrolizumab. A total of 65 patients were excluded from our analysis due to missing data (Fig. 1). The training cohort consisted of who 53 responders and 127 non-responders. (Table S3). The validation cohort consisted of 69 responders and 127 non-responders. There was a significant difference between the responders and non-responders with regard to the PD-L1 status ($p < 0{\cdot}001$).

3.2. Test design

The following test optimization minimum values were established for a test at 6 weeks: CA125 : 65 U/ml, CEA: 6g/L, Cyfra 21·1 : 4g/L, NSE: 20g/L, and SCC: 3·5g/L (Fig. S2). In the validation set, the specificity of CEA, NSE, SCC, Cyfra 21·1, and CA125 was 98·3% (95%CI: 90·9–100%), 96·5% (95%CI: 87·9–99·6%), 96·5% (95%CI: 88·1–99·6%), 91·8% (95%CI: 81·9–97·3%), and CA125 86·0% (95%CI: 74·2–93·7%) respectively (Table 2). Only the markers NSE and SCC showed a sensitivity below 20%. For SCC however, a small subset of patients with a squamous cell carcinoma showed an increase in the sensitivity of our test, without loss of specificity, from 6% (3·0–11·1%) to 15·4% (6·4 – 31·2%) (Table S6). The test accuracy for tumor marker to predict non-response was comparable between week 2 and 20 (Fig. 2).

Table 1 - Patient characteristics of the full cohort. All patients described have at least one baseline measurement and one follow-up measurement between weeks 2-20.

	TRAINING			VALIDATION				
	Non-responders (PD)	Responders (PR & SD)		Non-responders (PD)	Responders (PR & SD)		Total	
	N=127	N=53	p-value	N=127	N=69	p-value	N=376	p-value
Patient								
Male sex – no. (%)	75 (59.1)	27(50.9)	0.317	65(51.1)	35(50.7)	0.951	202(53.7)	0.273
Age (years) – mean (SD)	62.8(SD: 10.727)	64.3(SD: 8.190)	0.375	62.9 (SD: 8.9)	62.1 (SD: 8.9)	0.513	62.9 (SD: 9.5)	0.542
Smoking – no. (%)	25(19.7)	3(5.7)	0.010	17(13.4)	4(5.8)	0.249	49(13.0)	0.384
Pack years – mean (SD)	31.6(SD: 18.85)	35.5(SD: 19.3)	0.256	36.3 (SD: 19.5)	34.2 (SD: 17.5)	0.486	34.3 (SD 18.9)	0.240
WHO ≥ 2– no.(%)	20(15.7)	3(5.7)	0.047	13(10.2)	3(4.3)	0.156	39(10.4)	0.148
Tumor characteristics								
Pathology– no.(%)								
Adenocarcinoma	94(74.0)	33(62.3)		84(66.1)	39(56.5)		250(66.5)	
Squamous	22(17.3)	13(24.5)	0.286	24(18.9)	18(26.0)	0.285	77(20.5)	0.176
Other	11(8.7)	7(13.2)		19	12(17.4)		49(13.0)	
Mutations– no.(%)								
EGFR positive	4(3.1)	0	0.182	7(15.0)	2(2.0)	0.297	13(3.5)	0.192
KRAS positive	37(29.1)	15(28.3)	0.827	41(32.3)	25(36.2)	0.923	118(31.4)	0.020
BRAF	5(3.9)	2(1.6)	0.813	4(3.1)	1(1.4)	0.440	12(3.2)	0.405
ALK	0	0	-	2(1.6)	1(1.4)	0.898	3(0.8)	0.053
PD-L1 – no.(%) [1]								
Unknown	65(51.2)	20(37.7)	0.358	55(43.3)	28(40.6)	0.022	167 (44.4)	0.129
PD-L1 <1%	38(61.3)	17(51.5)	0.025	40(55.6)	14 (33.3)	<0.001	109 (52.2)	<0.001
PD-L1 >1%	24(38.7)	16(48.5)		32(44.4)	28 (66.7)		100 (47.8)	0.536
PD-L1 >50%	8(12.9)	10(18.9)	-	17(23.6)	24(57.1)		59 (28.2)	
Brain Metastasis – no.(%)	26(20.5)	12(22.6)	0.745	25(19.7)	11(15.9)	0.578	74 (19.6)	0.536

Table 1 - Patient characteristics of the full cohort. All patients described have at least one baseline measurement and one follow-up measurement between weeks 2-20. *(Continued)*

	TRAINING			VALIDATION				
	Non-responders (PD)	Responders (PR & SD)		Non-responders (PD)	Responders (PR & SD)		Total	
	N=127	N=53	p-value	N=127	N=69	p-value	N=376	p-value
Treatment								
Nivolumab	127(100)	53(100)	-	104(81.9)	51(73.9)	0.190	335 (89.1)	<0.001
Pembrolizumab	0	0	0.465	23(18.1)	18 (26.1)	0.400	41 (10.9)	
Line of treatment – no.(%)								
1st line	3(2.4)	0		2(1.6)	3(4.3)		8 (2.1)	
2nd line	90(70.9)	40(75.5)	-	99(78.0)	54(78.3)	-	283(75.3)	0.304
≥ 2nd line	33(26.9)	12(22.6)		26(20.5)	11(15.9)		82(21.8)	
Deceased after 6 months	74(58.3)	0		59(46.5)	0		133(35.4)	0.026
Comorbidities								
Auto Immune Disease – no.(%)	6(4.7)	0	0.106	8(6.3)	6(8.7)	0.538	20 (5.3)	0.119

[1] *Percentages shown are based on total known PD-L1 scores.*
Abbreviations: N: Number of patients; SD: Standard Deviation; no.: Number of patients, ECOG performance-status score: European Cooperative Oncology Group performance status score, this is a score ranging from 0 to 5, where 0 indicates no symptom, 1 indicates mild symptoms and above 1 indicates greater disability; EGFR: Epidermal Growth Factor Receptor; KRAS: Kirsten rat sarcoma viral oncogene; BRAF: v-raf murine sarcoma viral oncogene homolog B *; ALK: Anaplastic Lymphoma Kinase; PD-L1: Programmed death ligand 1.*

Table 2 - Results of the training cohort and the tests validated in the validation cohort.

Test	Minimum value	Results training			Results validation		
Single marker test							
		Sensitivity	**Specificity**	**PPV**	**Sensitivity**	**Specificity**	**PPV**
CA125	**65 U/ml**	21.7% (13.4-32.1%)	97.9% (88.7-100%)	94.7% (74.0-99.9%)	25.0% (16.0-35.9%)	86.0% (74.2-93.7%)	71.4% (51.3-86.8%)
CEA	**6 µg/L**	19.8% (12.0-29.8%)	100% (92.8-100%)	100% (80.5 – 100%)	20.7% (12.6-31.1%)	98.3% (90.9-100%)	94.4 % (72.7 – 99.9%)
	12 µg/L	15.1% (8.3-24.5%)	100% (92.8-100%)	100% (75.3-100%)	16.1% (9.1-25.5%)	98.4% (91.2-100%)	93.3% (68.1-99.8%)
Cyfra 21.1	**4 µg/L**	31.8% (22.1-42.8%)	100% (92.8-100%)	100% (87.2-100%)	34.5% (24.6-45.5%)	91.8% (81.9-97.3%)	85.7% (69.7 – 95.2%)
	8 µg/L	23.5% (15.0-34.0%)	100% (92.8-100%)	100% (83.2-100%)	25.3% (16.6-35.8%)	95.1% (86.3-99.0%)	88.0% (68.8-97.5%)
NSE	**20 µg/L**	13.3% (6.8-22.5%)	100% (92.1-100%)	100% (71.5-100%)	12.7% (6.2-22.1%)	96.5% (87.9-99.6%)	83.3% (51.6 – 97.9%)
	40 µg/L	4.8% (1.3-11.9%)	100% (92.1-100%)	100% (39.8-100%)	8.9% (3.6-17.4%)	96.5% (88.0-99.6%)	77.8% (40.0-97.2%)
SCC	**3.5 µg/L**	9.6% (4.3-18.1%)	97.9% (88.9-100%)	88.9 (51.8 – 99.7%)	2.4% (0.3-8.5%)	96.5% (88.1-99.6%)	50% (6.7-93%)
Combinations							
		Sensitivity	**Specificity**	**PPV**	**Sensitivity**	**Specificity**	**PPV**
Cyfra 21.1 OR CEA	**Cyfra: 4µg/L CEA: 6 µg/L**	38.4% (28.1-49.5%)	100% (92.8-100%)	100% (89.4-100%)	40.2% (29.9-51.3%)	91.8% (81.9-97.3%)	81.8% (64.5 – 93.0%)
Cyfra 21.1 OR CEA OR NSE	**Cyfra: 4µg/L CEA: 6 µg/L NSE: 20 µg/L**	38.4% (28.1-49.5%)	100% (92.8-100%)	100% (89.4-100%)	43.7% (33.1-54.7%)	91.9% (82.2-97.3%)	82.7% (66.3 – 93.4%)
Cyfra 21.1 OR CEA	**Cyfra: 8 µg/L CEA: 12 µg/L**	30.2% (20.8-41.1%)	100% (92.8-100%)	100% (86.8-100%)	28.7 % (19.5 – 39.43%)	95.1% (86.3-99.0%)	89.3 % (71.8-97.8%)
Cyfra 21.1 OR CEA OR NSE	**Cyfra: 8 µg/L CEA: 12 µg/L NSE: 40 µg/L**	30.2% (20.8-41.1%)	100% (92.8-100%)	100% (86.8-100%)	32.2 % (22.6 – 43.1%)	95.2% (86.5 – 99.0%)	90.3 % (40.7 – 59.3%)

Each marker test was defined as positive for the prediction of non-response if the value of that tumor marker met two criteria: (i) elevation of 50% compared to baseline and (ii) above a minimum value(second column in the table). All results, such as sensitivity, are given as a percentage (95% confidence interval). PVV: positive predicted value. µg/L: microgram per liter; U/ml: Units per milliliter.

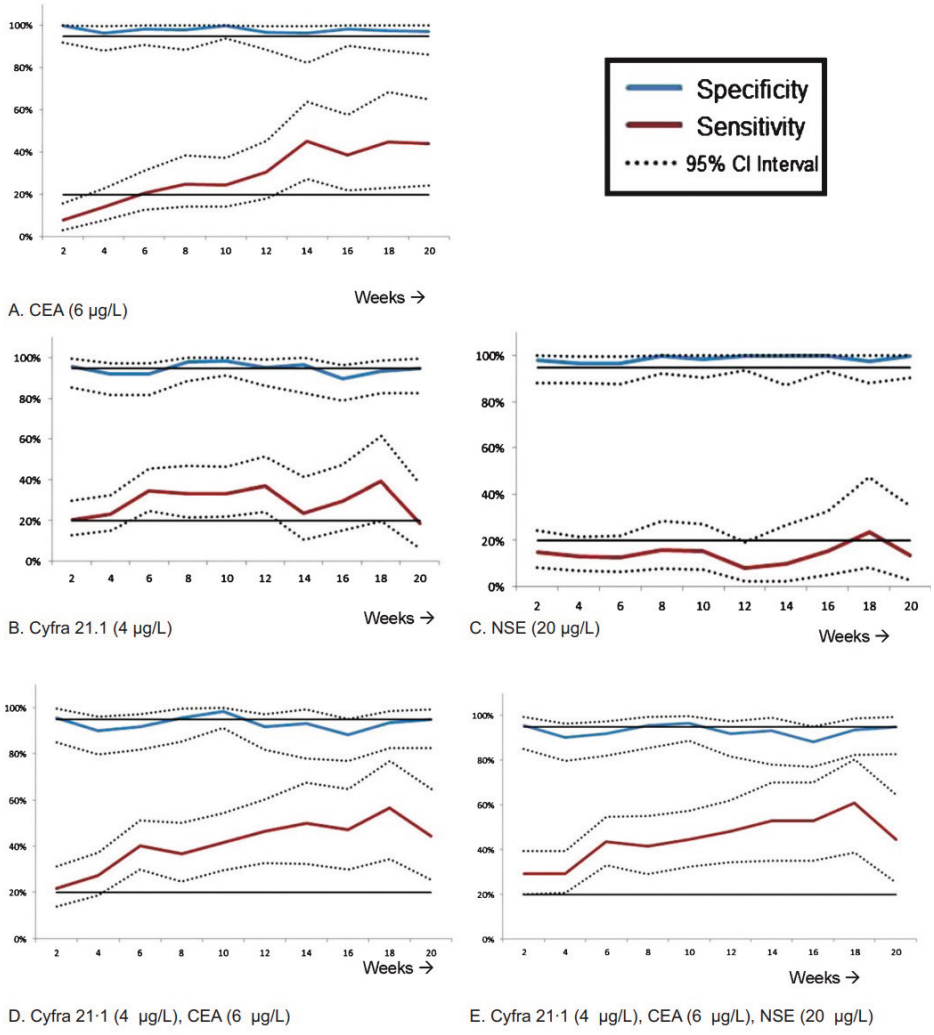


Figure 2 - Test characteristics for week 2–20 in the validation cohort, shown as sensitivity and specificity per week.
The horizontal as indicates the tests done every other week. Every time point displayed is that week and the week before (i.e. the time period for week 2 is week 1-2). If there was more than one measurement in this time period, the latest measurement was taken. The combination of markers were considered positive if at least one of the tumor markers had a positive test result. The two, straight lines indicate 20% and 95% respectively and are choosen for improved visibility. *µg*/L: microgram per liter; U/ml: Units per milliliter.

In the validation set the combination of Cyfra 21.1, CEA with or without NSE showed a specificity of 91.9% and 91.8% respectively and a sensitivity of 40.2–43.7% (Table 2). With these results, we decided to also validate a more stringent test by doubling the minimum value (Table S2). The specificity increased to 95.1% (86.3– 99%) with NSE and 95.2% (86.5–99.0%) without NSE at the cost of a lower sensitivity (28.7–32.2%). The results of the 23 performed tests in the training set with different markers can be found in the

supplemental material (Table S2 and Fig. S5). Also for the combination tests, we studied the same minimum value in other serial time points during treatment (week 2– week 20). The diagnostic performance of combined tumor biomarker tests for different follow-up are presented in Fig. 2.

3.3. False positive analysis

With the test at 6 weeks there were in total 13 patients (3.5% of the total cohort) with a false-positive result (Table S5). In the total cohort, 9 patients showed a false-positive result for CA125. In five out of these 13 patients Cyfra 21.1 showed a false-positive result. Two out of these five patients had a PR at 6 months from which one patient actually had a pseudo progression at 6 weeks, as was confirmed with a CT-scan. The other patient had an active hyperthyroidism at the start of treatment, which might explain the increases of the tumor markers (up to 7012% for SCC). For the other two patients, no specific explanation was found (TableS5)

3.4. Survival outcome

The median OS and PFS for the patients in the validation set were 363 days (95% CI 317–409 days) and 130 days (95%CI 98–162 days) respectively (Fig. 3). The median OS and PFS of patients depicted as non-responsive versus responsive were 153 days (95% CI 139–167) and 58 days (95%CI 46–70 days) versus 450 days (95% CI 347–553 days) and 237 days (95%CI 185–289) respectively ($p < 0.001$).

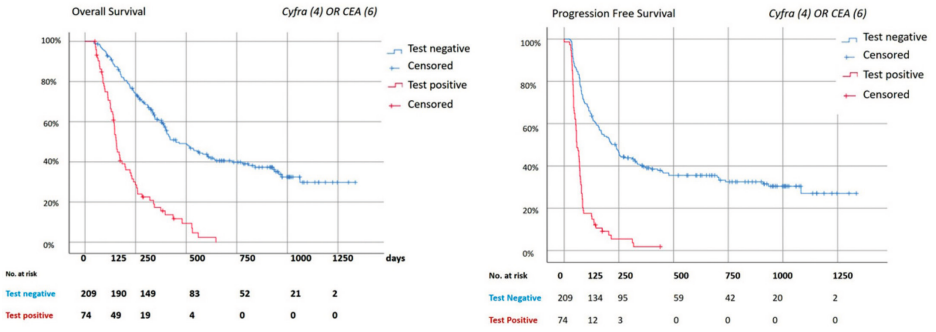


Figure 3 - Survival analysis.

Kaplan Meier analysis for the combination of Cyfra (4µg/L) and CEA (6µg/L). The combination of markers were considered positive if at least one of the tumor markers had a positive test result. All the analysis were done with the patients who had a test at week 6, as described in table S3. The median follow-up time was 322 days (IQR: 157–606 days). Date of last follow-up was 28th of January, 2019. A: Overall Survival. Median overall survival: 363 days (95% CI 317–409 days). Median OS negative test: 450 days (95% CI 347–553 days); Positive test: 153 days (95% CI 139–167 days). Log Rank (Mantel-Cox): $p < 0.001$. B: Progression Free Survival. Median progression free survival (PFS) 130 days (95% CI 98–162days). Median PFS negative test: 237 days (95% CI 185–289). Median PFS positive test: 58 days (95% CI 46–70 days). Log rank (Mantel-Cox) $p < 0.001$.

3.5. Pembrolizumab first line

In a small cohort of 31 patients who received pembrolizumab as first line treatment, an analysis was done. Results were comparable (Table S7 and S8).

4. Discussion

With the introduction of immunotherapy for metastasized NSCLC and its limited efficacy more tailored treatment strategies are needed. As far as we know, this is the first study that describes how to use liquid biopsy data for early treatment decisions in patients without clinical benefit from immunotherapy. In this prospective, observational study cohort a serum tumor markers panel was clinically validated as an early response tool that accurately predicts non-response to immunotherapy. These results indicate that serum tumor markers can be used to identify patients in which treatment can be discontinued early safely because it is ineffective. This potentially results in lower risk of side effect, lower costs, and allows alternative treatment options, while the patient is still in a good condition.

A commonly used and investigated liquid biopsy biomarker is ctDNA, which derives from normal physiological tissue remodeling events, necrosis and/or apoptosis of cancer cells [290-292]. The study of Goldberg et al. [250] showed the dynamics of ctDNA during immunotherapy treatment. In this study all patients with confirmed PR showed a ctDNA drop of > 50%, suggesting this is a helpful tool for monitoring response during treatment, although the dynamics of ctDNA in patients with progressive disease were more dynamic. Also, the strictly individual patterns of mutations complicate implementation in general practice of ctDNA-based response assessment and moreover a technical standardization for ctDNA is not yet available [292]. On the contrary, tumor markers are widely used, measured and implemented in clinical practice for years, making them a good alternative as a potential liquid biopsy. There is some literature about the role of serum tumor markers to assess efficacy of systemic treatment of NSCLC. Noonan et al. [293] showed in their study in patients with a targetable driver mutation in a smaller analysis that 59% of these patients, mostly responders, showed an increase right after start of their treatment. In the majority of patients marker concentration in plasma normalized to the baseline value during treatment. This shows the possible relation between tumor response and the measured markers. Furthermore, in the recent article of Dal Bello et al. [288], they measured CEA, Cyfra 21.1, and NSE at multiple time points. Their aim was to use these tumor markers for the monitoring of response (PR and SD). With their designed test, a decrease of 20%, identified responders. Interestingly enough, they also found that their test yielded similar results in the first and the fourth cycle of nivolumab. However, the study did not provide a tool to use these markers in optimizing treatment strategies, neither did other studies [282]. Therefore, our dataset with more than 400 patients and serial tumor marker data is contributing to the development of a clinical tool.

The requirements of a tumor marker test for early treatment decisions are depending on the clinical application. The current standard of care is to treat all patients with immunotherapy with or without chemotherapy, depending on the PD-L1 status [294]. A high specificity is required to prevent discontinuation of treatment in patients with a potential benefit. This approach, was also advocated in a study on the usage of an electronic nose. De Vries et al. [260] were able to identify 24% of the non-responders at baseline by exhaled breath analysis. On the other hand, the test should have an added value. Therefore the percentage of patients who will not respond and have a positive test, in other words sensitivity, must be contributing to current standards (e.g. radiological response and clinical assessment). In this study, we aimed to find the right balance of

these factors per individual and in combination of different tumor markers. Although not all individual markers showed a sufficient sensitivity (NSE and SCC), combining markers increases the sensitivity thereby optimizing its clinical utility.

Decisions during treatment are depending on radiological assessment and clinical performance; often treatment is continued despite the fact that the condition of the patient is deteriorating. In our cohort, the specificity of the CT-scan was 96.8% [24], but the therapy is only discontinued following confirmed radiological progression or in case of clinical deterioration. Besides, often no measurable lesions are available for radiological response assessment. Having established tumor markers as robust tool to establish non-responsiveness, we postulate that tumor markers can improve early treatment decision making. In the future, combining radiology and tumor markers, together with assessment of the clinical condition, will likely improve overall test accuracy.

Spikes are a well-known phenomenon seen in liquid biomarker research in patients with response [295, 296], which is also shown for immunotherapy [250]. However, in our study, spikes were not common (Fig. S6). Nevertheless, there were patients with a false-positive result. Renal failure, liver failure or (other) lung diseases are known causes for multiple different elevated tumor markers [297]. We did not see this in our cohorts, maybe due to the selection criteria of immunotherapy. What is more commonly known is the lack of accuracy of CA125. False positive results are often present in case of a serositis [298]. There was one patient with an active thyroiditis (Table S5A, Patient A) with extremely elevated tumor marker levels. We are not aware of data supporting the correlation of thyroiditis with tumor marker elevation. Our findings suggest that tumor marker tests should be treated with caution in case of an active thyroiditis.

A strength of this study is the homogeneous patient population with mainly ≥ 2nd line NSCLC patients treated with single agent immune checkpoint inhibitors and the use of independent training and validation cohort, which makes it an robust analysis. However, there are a few limitations of our study to be considered. Firstly, therapeutic options are rapidly changing and patients are currently treated with immunotherapy as their first line of treatment[293]. We included a small analysis with first-line pembrolizumab patients, in order to assess the utility of the tests in the current standard of treatment. Mature data of a small cohort (n = 23) of patients, treated with first line pembrolizumab, was available and results were comparable. However, larger validation studies are warranted. Secondly, there training cohort consisted of patients who were treated with nivolumab only. We are uncertain if this might cause bias. Pembrolizumab and nivolumab are both PD-1 inhibitors and the validation results in the pembrolizumab cohort were comparable. Thirdly, in our study, we validated a minimum value instead of a percentage. These minimum values and reference values differ between different hospitals. However, all of the chosen minimum values are more or less multiplied by a round number between one and three (Fig. S2), allowing a relatively easy validation of these tests in other hospitals.

All in all, in this study we designed and validated tests with single and multiple serum based tumor markers for the early prediction of non-response. Based on our results, serum tumor marker based response monitoring can be used for clinical decision making in NSCLC treated with immunotherapy. Future studies are required to determine the added value in clinical practice.

Supplemental data

The supplemental material can be found at:

Contents included in this thesis:
Figure S1 – Schematic view
Figure S2 – Cut-off check training set
Figure S3 – Week 2-20 for the training cohort
Figure S4 – Week 2-20 for the validation cohort
Table S1 – Considerations for the development of a marker test
Table S5 – False-positives for test at week 6.
Table S6 – Sub analysis Squamous Cell Carcinoma and SCC marker
Table S7 – Patient characteristics of the first line pembrolizumab cohort
Table S8 – Analyses first line pembrolizumab cohort

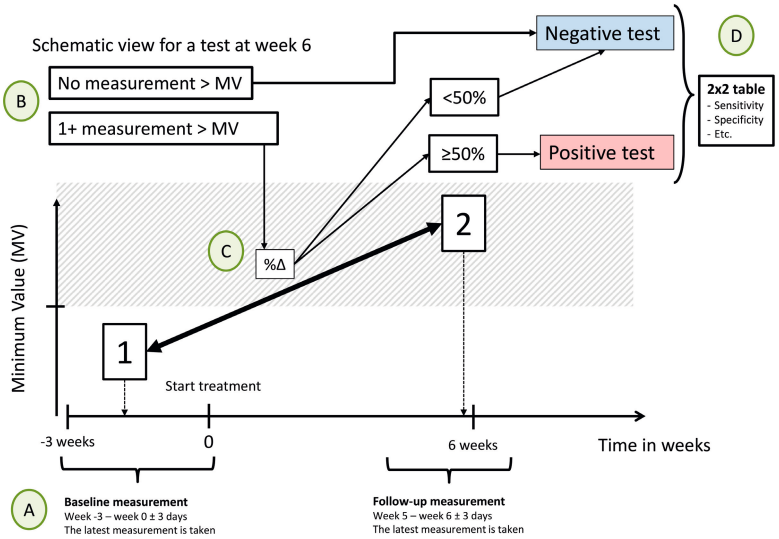


Figure S1 - Schematic view of the application
A: The application checks if a patient has measurements (in figure referred to as 1 and 2) within the given timeframes. If there is more than one measurement in a given baseline or follow-up timeframe, the last measurement will be used. If measurements within one of the two timeframes are missing, the results are obviously defined as non-conclusive (and therefore will not be add to the 2x2 cross table). **B:** The program checks if at least one of these measurement is above the minimum value (MV). The minimum value is a threshold used to avoid background noise from non-relevant increases of a marker. If both values are below this threshold, the test will be defined as negative. **C:** If abovementioned criteria are fulfilled, than the difference in percentage (%Δ = (measurement 2 – measurement1) / measurement1 x 100%) will be calculated**.** In our study, and in alignment with the results from the ReMarker analysis, the test was deemed to be positive when the marker in a patient had increased with 50%. **D:** The outcome of the test of a patient is used in a 2x2 cross table and compared to clinical data.

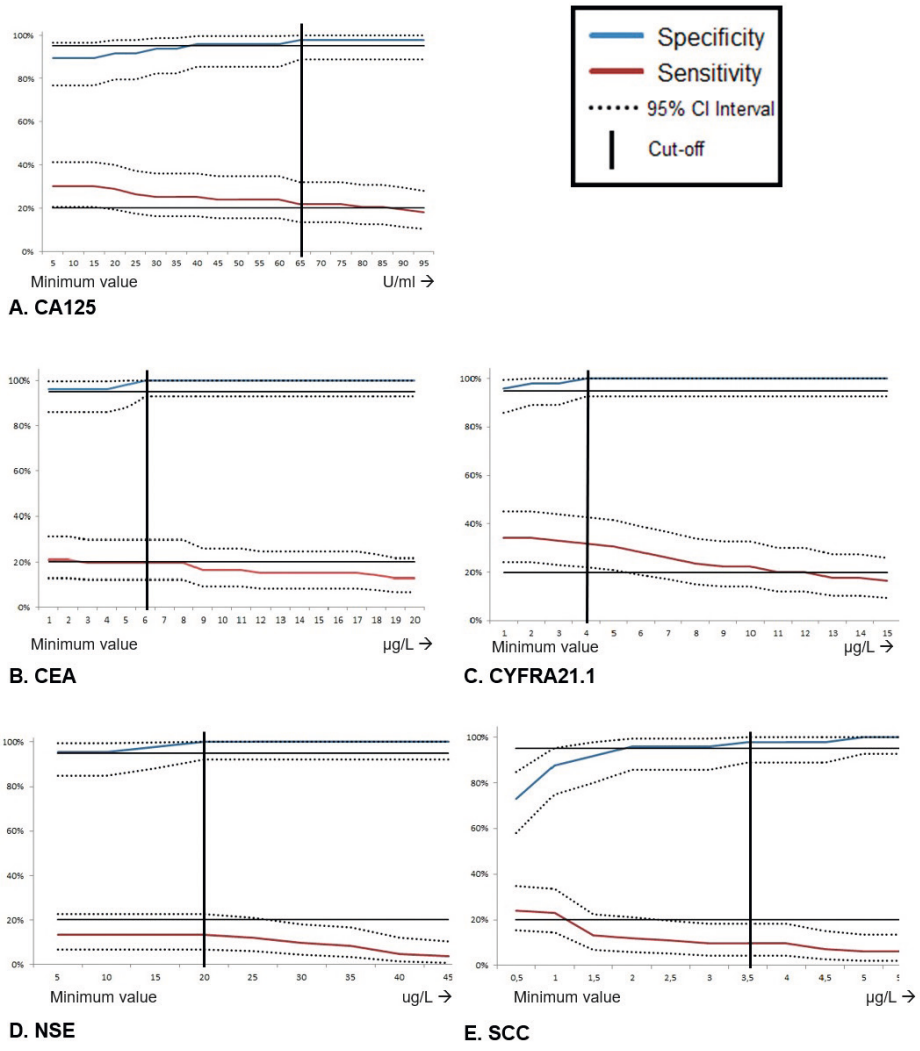


Figure S2 – The optimal minimum value at week 6 in the training cohort
µg/L: microgram per liter; U/ml: Units per milliliter.

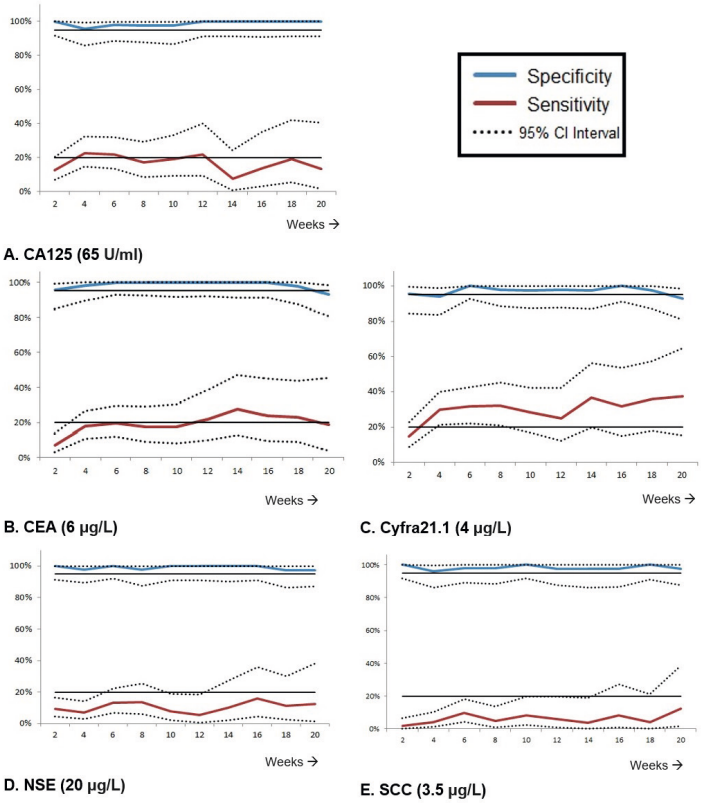


Figure S3 – Characteristics of the serum tumor marker cut-off values for week 2 – 20 in the training cohort, shown as specificity and sensitivity.

The horizontal axis indicates the tests done every other week. Every time point displayed is that week and the week before (i.e. the time period for week 2 is week 1-2). If there was more than one measurement in this time period, the latest measurement was taken. The two, straight lines indicate 20% and 95% respectively and are choosen for improved visibility.µg/L: microgram per liter; U/ml: Units per milliliter.

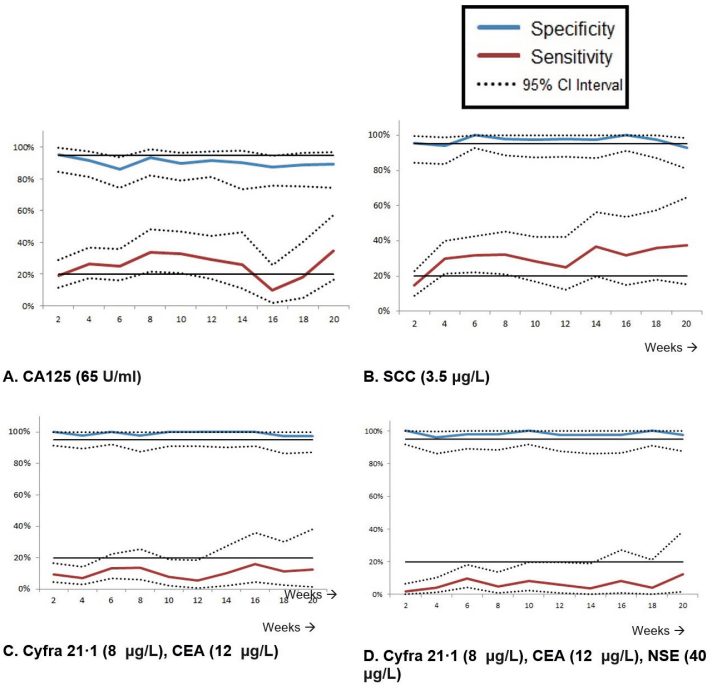


Figure S4 – Results of the test characteristics of tumor markers in the validation set, shown as sensitivity and specificity per week.

Notes: The combination of markers were considered positive if at least one of the tumor markers had a positive test result. The horizontal as indicates the tests done every other week. Every time point displayed is that week and the week before (i.e. the time period for week 2 is week 1-2). If there was more than one measurement in this time period, the latest measurement was taken. The two, straight lines indicate 20% and 95% respectively and are choosen for improved visibility. µg/L: microgram per liter; U/ml: Units per milliliter.

Table S1 - Considerations for the development of a marker test.

Considerations		Single marker test
1	Easy to use in clinical practice	
2	>50% increase	>50% increase compared to baseline Based on earlier BReC plot analysis [14] Considered easy-to-calculate
3	Minimum value	The minimum value is used instead of the reference value of a marker. Reference values in the clinical chemistry are based on the fact that 95% of all test results lay below the reference value, measured in the healthy population [294]. However, the aim is to identify non-responders in a group of lung cancer patients. Therefore, we introduced a new value, the minimum value. The role of the minimum value criterion is applied to exclude patients with small biomarker increases at low concentrations that results in large relative increases thereby reducing the effect of (pre-) analytical and biological "background noise" At least one of the measurements should be above the minimum value (as described in figure S1)
4	Optimal minimum value	The optimal minimum value per marker was determined by calculating the specificity and sensitivity for predicting the clinical endpoint per minimum value (figure S2)
5	Specificity > 97,5%	The test was developed as an early treatment decision tool and should be very accurate in detecting non-responsiveness (to safely discontinue treatment) Minimum values yielding a specificity of ≥97,5% in the training set per individual markers were considered a good cut-off, in order to increase the likelihood to achieve a specificity of >95% in the validation set
6	Sensitivity >20%	The test should have added value over the current standard, i.e. decisions based on radiological and clinical assessment. Observation in the training set is that about 20% of the patients discontinue treatment within 6 weeks based on radiological assessment and/or experienced clinical deterioration. Therefore a sensitivity >20% was considered a good cut-off. A combination of tumor markers was assumed to increase in sensitivity, therefore for the single biomarkers a sensitivity less than 20% was accepted in the training cohort
		A test with a combination of markers
7	Combination of single marker tests	Two single marker tests with the previous described characteristics are used of a combination test
8	At least one test is positive	The combination of markers were considered positive if at least one of the tumor markers had a positive test result, defined by an increase of the marker concentration according to abovementioned criteria When at least one of the markers is positive, the sensitivity of the test is likely to increase.
9	Two times the minimum value	Considered easy to calculate The combination of markers may lead to a decrease in specificity, since all patients with a false-positive tests are added together in the combination test, leading to more false-positive results.

Table S5.A - False positives in the final test (Cyfra/CEA/NSE)

ID	Diagnosis	Treatment	Response	Cohort	CA125	CEA	Cyfra	NSE	SCC	Explanation
A	Ad	Nivo	PR	Val	424%	225%	218%	89%	7012%	Hyperthyroidism / thyroiditis
B	Ad	Pembro	PR	Val	205%	-	61%	-	-	Pseudo progression
C	Sq	Pembro	SD	Val	-	-	67%	-	-	"Progressive SD"
D	Ad	Pembro	SD	Val	-	-	81%	-	-	Small amount of pleural fluid
E	Ad	Nivo	SD	Val	86%	-	170%	90%	-	Nephrodrain, relatively normal kidney function (GFR 43)

Table S5.B - False positives in the remaining markers (CA125 and SCC)

ID	Diagnosis	Treatment	Response	Cohort	CA125	CEA	Cyfra	NSE	SCC	Explanation
F	Ad	Nivo	SD	Val	151%	-	-	-	-	"Progressive SD"
G	Sq	Nivo	SD	Val	51%	-	-	-	-	None
H	Ad	Nivo	SD	Train	59%	-	-	-	-	None, although suffering from brain infarction
I	Ad	Nivo	SD	Val	111%	-	-	-	-	None
J	Ad	Nivo	SD	Val	70%	-	-	-	-	Skin rash grade 2
K	Ad	Pembro	PR	Val	100%	-	-	-	-	Pseudo progression
L	Ad	Nivo	SD	Val	-	-	-	-	63%	None
M	Sq	Nivo	SD	Train	-	-	-	-	92%	Progressive pleural metastasis.

Table S5 – False positives.

A. The false positive results in the final test for Cyfra/CEA/NSE. For all the false-positive tests in the final test at 6 weeks after start, which was defined as a responder classified by non-responder with our test, with a possible explanation. B. The false positive results for the remaining markers. ID: Identification number of a patient; Response: the response after six months according RECIST criteria, with SD: Stable disease and PR: partial response ; PFS: Progression free survival in days; Ad: Adenocarcinoma; Sq: Squamous Cell Carcinoma; Nivo: nivolumab; Pembro: Pembrolizumab; Val: Validation cohort; Train: Training cohort.

Table S6 – Sub analysis SCC

Minimum value SCC		Specificity	Sensitivity	Positive predicted value
0.0 µg/L	All pathology	76.4% (67.0-83.9%)	22.9% (16.9-30.2%)	60.3% (47.2-72.2%)
	Squamous	80.8% (60.0-92.7%)	41.0% (26.0-57.8%)	76.2% (52.5-90.9%)
2.0 µg/L	All pathology	95.3% (88.8-98.3%)	9.6% (5.8-15.4%)	76.2% (52.4-90.9%)
	Squamous	88.5% (68.7-97.0%)	23.1% (11.7-39.7%)	75.0% (42.8-93.3%)
3.5 µg/L	All pathology	97.2% (91.3-99.3%)	6% (3-11.1%)	76.9% (46.0-93.8%)
	Squamous	96.2% (78.4-99.8%)	15.4% (6.4-31.2%)	85.7% (42.0-99.2%)

Sub analysis of the full cohort versus squamous cell only. µg/L: microgram per liter.

Table S7 – Patient characteristics pembrolizumab first line

Pembrolizumab first line	Non-responders (PD)	Responders (PR & SD)	
	N=8	N=23	*P-value*
Patient			
Male sex – no. (%)	4	7	*0.319*
Age (years) – mean (SD)	57.1 (SD: 6.9)	65.5 (SD: 9.4)	*0.029*
Smoking (never) – no. (%)	0	0	-
Pack years – mean (SD)	35.4 (SD 27.9)	27.6 (SD: 13.2)	*0.279*
WHO ≥ 2– no.(%)	3	1	*0.018*
Tumor characteristics			
Adenocarcinoma	6	17	*0.646*
Squamous	0	2	*0.155*
Other	2	4	-
KRAS positive	3	16	*0.115*
PD-L1 >50%	8	23	
Brain Metastasis – no.(%)	2	1	

Abbreviations: N: Number; SD: Standard Deviation; no.: Number of patients, WHO: performance-status score: World Health Organization performance status score, this is a score ranging from 0 to 5, where 0 indicates no symptom, 1 indicates mild symptoms and above 1 indicates greater disability; KRAS: Kirsten rat sarcoma viral oncogene; PD-L1: Programmed death ligand 1.

Table S8 – Results of tumor marker test at 6 weeks (5-7 weeks) for Pembrolizumab monotherapy.

Setting	Sensitivity (95%-CI)	Specificity (95%-CI)	Positive Predicted Value (95%-CI)
CEA 6 µg/L OR Cyfra 4 µg/L	25,0% (4.5-64.4%)	95,6% (76.0-99.8%)	66,7% (12.5-98.2%)
CEA 6 µg/L OR Cyfra 4 µg/L OR NSE 20 µg/L	25,0% (4.5-64.4%)	95,6% (76.0-99.8%)	66,7% (12.5-98.2%)
Cyfra 10 µg/L OR CEA 10 µg/L	20% (3-56%)	96% (80-100%)	66% (15-46%)
Cyfra 10 µg/L OR CEA 10 µg/L OR NSE 20 µg/L	20% (3-56%)	96% (80-100%)	66% (15-46%)

A total of 62 patients were treated with pembrolizumab first line. From 27 patients there was no test: 4 patients were non-evaluable, 7 patients were lost to follow up (treatment continuation in another hospital), 20 patients missed either the baseline or follow-up measurement. Data from the remaining 31 patients was used for the analysis. The same criteria were used as in the manuscript. Patient characteristics of this cohort can be found in table S.

9 Modelling Diagnostic Strategies to manage toxic adverse-events following cancer immunotherapy

F. van Delft*, M. Muller*, R. Langerak, H. Koffijberg, V. Retèl, D. van den Broek, M. IJzerman.

Med Decis Making. 2021;41(6):693-705.

*Contributed equally

Abstract

Background: Although immunotherapy (IMT) provides significant survival benefits in selected patients, approximately 10% of patients experience (serious) immune-related adverse events (irAEs). The early detection of adverse events will prevent irAEs from progressing to severe stages, and routine testing for irAEs has become common practice. Because a positive test outcome might indicate a clinically manifesting irAE that requires treatment to (temporarily) discontinue, the occurrence of false-positive test outcomes is expected to negatively affect treatment outcomes. This study explores how the UPPAAL modeling environment can be used to assess the impact of test accuracy (i.e., test sensitivity and specificity), on the probability of patients entering palliative care within 11 IMT cycles.
Methods:A timed automata-based model was constructed using real-world data and expert consultation. Model calibration was performed using data from 248 non–small-cell lung cancer patients treated with nivolumab. A scenario analysis was performed to evaluate the effect of changes in test accuracy on the probability of patients transitioning to palliative care.
Results: The constructed model was used to estimate the cumulative probabilities for the patients' transition to palliative care, which were found to match real-world clinical observations after model calibration. The scenario analysis showed that the specificity of laboratory tests for routine monitoring has a strong effect on the probability of patients transitioning to palliative care, whereas the effect of test sensitivity was limited.
Conclusion: We have obtained interesting insights by simulating a care pathway and disease progression using UPPAAL. The scenario analysis indicates that an increase in test specificity results in decreased discontinuation of treatment due to suspicion of irAEs, through a reduction of false-positive test outcomes.

Background

Non–small-cell lung cancer (NSCLC) is associated with significant mortality. The incidence of lung cancer is estimated to be 11.6% of all new cancer diagnoses worldwide and is considered the leading cause of cancer-related mortality [299]. Until September 2018, in the Netherlands, the first-line treatment in the metastatic disease setting (stage IV NSCLC) was chemotherapy with platinum doublets in patients without a targetable mutation, and patients presenting with a targetable mutation (e.g., epidermal growth factor receptor or anaplastic lymphoma kinase mutation) received a targeted therapy (e.g., erlotinib, crizotinib, or gefitinib).[13] Since then, the treatment landscape has fundamentally changed with the introduction of immunotherapy, which now has become a standard treatment. Initially, nivolumab and pembrolizumab were approved in the second-line setting, but, since 2018, immunotherapy based on PD-L1 expression with or without added doublet chemotherapy has been the standard first-line treatment. Clinical studies have shown that about 20% to 40% of patients respond to immunotherapy, with substantially prolonged survival benefit. Despite the clinical benefits of the use of immunotherapy treatments, immunotherapy is also known to be associated with immunogenic reactions that severely affect treatment schedules and outcomes [300]. To manage immunotherapy-related adverse events (irAEs), suspicion of irAEs is determined using a standardized set of blood tests. However, no clear guidance is available on aspects such as the frequency of tests, required test specifications, or interpretation of results. This might result in suboptimal diagnostics and outcomes, in terms of survival [301].

At the Netherlands Cancer Institute, a diagnostic panel was implemented and routinely used at 2-weekly intervals. Test outcomes are used to aid clinical decision making on treatment continuation, to confirm the suspicion of an irAE, and to grade the severity of an irAE. The diagnostic panel aimed to detect irAEs in an early stage to improve clinical management and outcomes. The diagnostic kit covers a broad spectrum of blood markers. Some of these tests are used in a different setting (e.g., screening for irAE without clinical complaints) as compared with their original routine use, which could result in suboptimal performance in detecting irAEs. In addition, there are no data on the optimal frequency and use of diagnostic tests for irAEs. Hence, there is an interest in optimizing the diagnostic workflow to, for example, reduce unnecessary testing. Optimizing the test sequence using a prospective or retrospective study design to test different diagnostic setups would be unfeasible given the number of possible diagnostic strategies, time, and financial constraints. In such cases, it is possible to construct simulation models based on real-world data to evaluate diagnostic strategies aimed at the detection of irAEs on disease management and patient outcomes.

This study explores the use of UPPAAL to model diagnostic strategies. UPPAAL was developed by computer scientists at Uppsala University (Sweden) and Aalborg University (Denmark) [302]. The UPPAAL software uses a distributed modeling paradigm that allows for modeling a system using networks' of timed automata (TA). The UPPAAL tool enables modelers to construct networks of TA. These networks consist of a finite set of automata with real-valued clocks and constraints. Automata can be seen as a state in which a predefined process will be executed automatically (e.g., changing an integer according to a predefined function). Within the network, clock values increase with equal

speed, and clock values can be compared with integers to control transitions between automata[303]. Moreover, in these networks of TA, communication channels are used to allow for multiple types of synchronization signals that allow for communication between different automata in the network. Generally, TA models consist of multiple templates, with each template containing a network of automata used to model a specific function. The communication channels allow different templates to communicate with and influence each other. The ability to model substructures makes UPPAAL especially suited for modeling complex structures in which multiple agents have the ability to influence each other (e.g., a clinical pathway). In addition, UPPAAL provides extensive model-checking capabilities, which allow model developers to check the reachability of states or pathways. UPPAAL provides an environment that aids interdisciplinary communication (e.g., by using multiple templates to model subprocesses while using a graphical user interface). The heterogeneous treatment path and importance of event timing in the detection of irAEs requires a flexible modeling approach. In the field of health economics, patient-level Markov models, discrete event simulations (DES), or agent-based models are generally applied approaches when flexibility is required. A unique benefit of TA-based models over the currently preferred modeling approaches is the compositional model structure that allows modeled agents to interact with and influence each other through communication channels. This compositional nature of TA-based models makes them more flexible to adjust and also allows modeling of a continuous process that involves multiple decisions, as in the case of treatment of advanced cancers. A downside of TA-based models is the limitation in statistical distributions, which are limited to a uniform and exponential distribution in UPPAAL. However, in UPPAAL, there are workarounds that allow for the incorporation of other statistical distributions into the model[304].

Our research aims to use a TA-based routine to model the clinical-diagnostic pathway of irAEs and to populate and calibrate the model using real-world survival data. This model will then be used to demonstrate the feasibility of UPPAAL by evaluating different test scenarios with increasing diagnostic performance of a broad spectrum of irAEs tests. We hypothesize that a TA-based model created in UPPAAL will be versatile enough to capture the complexity of the clinical path and decision making.

Methods

Study Cohort

The model was developed using a cohort of patients treated with nivolumab through the compassionate use program and regular care, containing 248 patients. A description of the study cohort, response assessment, and safety assessment was published in Lung Cancer in 2017[24]. Of these patients, 133 were recruited through the compassionate use program, whereas 115 patients started treatment in regular care. All patients received at least 1 line of previous treatment (chemotherapy) before nivolumab. In August 2015, nivolumab was available through the compassionate access program by Bristol-Myers-Squibb in 8 different hospitals in the Netherlands (NCT02475382). In this program, patients who had received at least 1 previous line of anticancer treatment were eligible to receive nivolumab if they had a good clinical performance with no or mild symptoms (World Health Organization performance status 0–1) and had adequate lab values for blood markers specific for organ (dys)function (e.g., aspartate aminotransferase,

alanine aminotransferase, or creatinine)[247]. Data used in this study were limited to data acquired by the Netherlands Cancer Institute.

As part of routine care, patients were seen in the hospital every 2 weeks (wk), and laboratory tests were administered at baseline and every 2 wk thereafter. The laboratory assessment consisted of 30 blood tests including hematology, clinical chemistry, and hormonal measures, as depicted in Table 1.

Table 1 - An oversight of biomarkers included in the diagnostic panel.

Category	Measured biomarkers
Blood count	Hemoglobin, Hematocrit, Erythrocytes, MCV, Leukocytes, Neutrophil granulocytes, thrombocytes, cell differentiation
Liver function	Bilirubin, ALP, ASAT, ALAT, YGT, LDH
Clinical chemistry	CRP, Creatinine, GFR, Urea, sodium, potassium, phosphate, magnesium, glucose, total protein, albumin, calcium
Special chemistry	ACTH, Cortisol,

MCV, mean corpuscular volume; ALP, alkaline phosphatase: ASAT, aspartate aminotransferase; ALAT, alanine aminotransferase; YGT, gamma-glutamyl transferase; LDH, lactate dehydrogenase; CRP, C-reactive protein; GFR, glomerular filtration rate; ACTH, adrenocorticotropic hormone.

Disease progression was monitored through computed tomography imaging at 6 wk, 12 wk, 3 months (mo), 6 mo, 9 mo, 12 mo, and 15 mo after initiation of IMT. IMT was ceased in patients presenting with progressive disease. When any grade of irAE was clinically confirmed, patients were either withdrawn from IMT for the duration of a recovery period, which could take up to 5 wk, or IMT therapy was ceased definitively, and the patients proceeded to the next line of treatment. The NSCLC treatment landscape is heterogeneous. Therefore, in this model, the assumption is made that patients will transition to palliative chemotherapy after IMT is ceased definitively. During recovery, patients received appropriate treatment to recover from the incurred irAE. In practice, patients continued to the next line of therapy after discontinuation of IMT; in the model, we refer to this next line of treatment as “palliative care,” since only the IMT phase was incorporated in the model. The data used from this cohort included the time on treatment in weeks, the frequency, and the incidence of toxicities and progressive disease. All relevant irAEs incorporated in the model are described in Table 2.

Table 2 - *A* **Description of immune related Adverse Events included in the model, based on data from the Nivo chohort (n=248).** The primary function of the test lies in the detection of irAEs; however, these tests results are also part of the diagnostic process.

Adverse Event	Probability of developing irAE during IMT therapy (%)	Number of reported events in 248 patients (n)	Time to development (Median (Range, days))	Time between development and Grade 3-4 AE (weeks)	Symptoms	First indication / Laboratory assessment	First symptom	Confirmation of diagnosis	Course of treatment	Complications when not treated
Pneumonitis*	9.6	10	60 (10-120)	2-4	Stuffiness	Patient	Patient: stuffiness	CT-scan or Bronchoscopy	~4-8 weeks improvement of symptoms	Pulmonary fibrosis
Colitis*		7	45 (15-180)	4	Diarrhea		Patient: Diarrhea	Coloscopy	~2 weeks until symptom relief	Bowel perforation
Dermatitis*		6	30 (10-120)	NA	Often: itch		Patient: itch		~2-4 weeks using the proper ointment	
Arthritis *		2	60 (40-90)	12	Joint pain		Patient: Thickened joints, pain	Physical assessment by a medical specilist or possibly rheumatoid factor	2-4 weeks with pain medication	
Pancreatitis	1.2	3	180 (100-200)	2-4	Stomach ache	Patient and lab: amylase, lipase, liver functon	Patient: pain in abdomen. Lab: Increased amylase/ lipase			
Hepatitis	2.8	7	90 (30-150)	4	Jaundice, feeling ill	Lab: ASAT, ALAT, YGT, ALP, Bilirubin	biomarker assessment		~ 4 weeks until improvement of lab values	
Hypofysitis	1.6	4	120 (90-300)	2	Feeling ill	Lab: Cortisol, ACTH, Na, K	Lab, or patient: feels ill		hospitalization: 2-4 until improvement	
Pancytopenia	0.4	1	60 (NA)	1-8	Bleedings, bruising easily	Lab: Hemoglobin, white blood cell count and differentiation, platelets	biomarker assessment	Shortage / absence Thrombocytes, leukocytes, erythrocytes		Infections, severe bleeding, anemia
Diabetes	0.4	1	10 (NA)	4-8	Thirst, urinating, blurry vision	Lab: Glucose	biomarker assessment, sometimes: thirsty, fatigue	Glucose: elevated	Insulin	Coma in case of severely elevated glucose

ASAT, aspartate aminotransferase; ALAT, alanine aminotransferase; YGT, gamma-glutamyl transferase; ALP, alkaline phosphatase; ACTH, adrenocorticoptropic hormone; Na, sodium; K, calcium
* Grouped as one immune related adverse event (irAE) based on the assumption that these irAEs will manifest with clear physical symptoms, and are generally discoved by patients themselves.

Expert Consultations

A multidisciplinary team involving experts in computer science, medical oncology, laboratory medicine, epidemiology, and decision science was involved in the development of the model. During the model development phase, 5 meetings were arranged with the multidisciplinary team, in which the research questions, model structure, model inputs, and results were discussed.

Model Construction

Although TA-based models have been established in other fields, in decision science, TA-based models have rarely been used or published. The care pathway described in this study consists of 2 distinct events that are monitored independently during the treatment process (i.e., the development and detection of irAEs and the development and detection of disease progression). Markov or DES models use "events" or "timing of events" to dictate the flow of patients through the model. Most of these models are built around 1 decision and process (i.e., a flow of subsequent actions). However, these models are less able to model asynchronous, parallel processes with multiple decision points and events causing an interruption of a process at arbitrary moments. A DES model does provide a more flexible approach, and depending on software-specific abilities, a DES model should be able to reflect the 2 independent subroutines. However, it would require a more complex model structure for which competing risks are defined for each combination of events. When using a Markov model, short cycle times could be used to allow for the evaluation of events at each cycle. However, a Markov model is less able to capture complex pathways with time-varying probabilities. UPPAAL provides the ability to model independent processes asynchronously, while synchronization channels can be used to interrupt processes in subroutines when necessary. Other merits of using UPPAAL are the ability to create substructures that represent a specific aspect of the simulated pathway and its model-checking engine. The ability to model substructures aids interdisciplinary communication, since each substructure is assigned its own template in a graphical user interface. In addition, the model-checking engine enables model developers to check the reachability of each state or pathway. A high-level overview of the clinical pathway is depicted in Figure 1. This high-level overview was translated into 6 templates used to capture different parts of the clinical pathway. In our model, the IMT is stopped if the patient has received 11 cycles of IMT, the patient develops progressive disease, or treatment is ceased because of irAEs. Therefore, our model adopts a time horizon of 66 wk, that is, 11 treatment cycles with a duration of 6 wk per cycle. A comprehensive overview of all transition probabilities, time constraints, and the underlying data source is provided in Table 3.

Table 3 - Model parameters and a description of the source on which the model parameter was based.

Description	Value	Unit	Source
irAE Occurrence: Grouped (Pneumonitis, Colitis, Dermatitis, Arthritis) - Hepatitis - Hypophysitis - Pancreatitis - Pancytopenia - Diabetes			
Time treatment start - occurrence of irAE, lower bound	1 - 4 - 3 - 14 - 8 - 1	Weeks	Patient data
Time treatment start - occurrence irAE, upper bound	26 - 21 - 43 - 29 - 9 - 2	Weeks	Patient data
Growth period (G1 - .G2 or G2 - .G3–4)	2 - 2 - 1 - 1 - 2 - 2	Weeks	Expert opinion
Probability of irAE occurrence, cycle 1	0.096 - 0.028 - 0.016 - 0.012 - 0.004 - 0.004	Probability	Expert opinion
Probability of irAE occurrence, cycle 2	0.192 - 0.056 - 0.032 - 0.024 - 0.008 - 0.008		
Probability of irAE occurrence, cycle 2	0.288 - 0.084 - 0.048 - 0.036 - 0.008 - 0.008		
Probability of irAE occurrence, cycle 2 – cycle 11	0.400 - 0.100 - 0.060 - 0.050 - 0.016 - 0.016		
Disease progression			
Probability of disease progression; cycle 1	0.38	Probability	Patiënt data
Probability of disease progression; cycle 2	0.29		
Probability of disease progression; cycle 3	0.22		
Probability of disease progression; cycle 4	0.168		
Probability of disease progression; cycle 5	0.128		
Probability of disease progression; cycle 6	0.097		
Probability of disease progression; cycle 7	0.074		
Probability of disease progression; cycle 8	0.056		
Probability of disease progression; cycle 9	0.04		
Probability of disease progression; cycle 10	0.033		
Probability of disease progression; cycle 11	0.025		

Table 3 - Description of all model parameters and a description of the source on which the model parameter was based. *(Continued)*

Description	Value	Unit	Source
irAE Occurrence: Grouped (Pneumonitis, Colitis, Dermatitis, Arthritis) - Hepatitis - Hypophysitis - Pancreatitis - Pancytopenia - Diabetes			
Recovery			
Probability of recovery, grade 0 irAE (false-positive)	1	Probability	Expert opinion
Probability of recovery, grade 1 irAE, occurrence 1 - 2 - 3 - .3	1 - 0.9 - 0.8 - 0		
Probability of recovery, grade 2 irAE, occurrence 1 - 2 - 3 - .3	0.8 - 0.5 - 0 - 0		
Probability of recovery, grade 3–4 irAE, occurrence 1 - 2 - 3 - .3	0.5 - 0.2 - 0 - 0		
Probability of recovery, fast recovery	0.4		
Enter recovery after detection of a G2 or G3 irAE	0.5		
Duration recovery	5	Weeks	Expert opinion
Duration fast recovery	2	Weeks	Expert opinion
Maximum of irAEs allowed	6		Patient data
Diagnostic accuracy			
Sensitivity diagnostic path - applied to all 6 irAEs included	85	%	Model calibration
Specificity diagnostic path - applied to all 6 irAEs included	91		

irAE, immune-related adverse event.

The 6 templates in the model are referred to as "Protocol,""Patient,""Toxic,""Test,""Monitor," and "Progression Check." Each of these templates fulfills a specific role in the model. The Protocol template is built to indicate whether a patient should receive tests aimed at the detection of irAEs, move to a recovery state, or transition to palliative care. The Patient template keeps track of the physical state of a patient (e.g., the grade of irAE incurred). The Toxic template determines whether a patient will incur a certain irAE. For patients who would develop an irAE, the template is also used to determine the point in time at which the irAE will manifest itself, simulate the progression of the irAE to a more severe grade, and determine whether a patient will recover once the patient enters the recovery phase. The Test template simulates outcomes of tests aimed at the detection of irAEs. The Monitor template is used to log the time patients spend in palliative care, which

was chosen as the primary model outcome. Disease progression is modeled using the Progression Check template, in which the probability of disease progression is derived from patient data and decreases over time (Table 3). An extensive model description was published in an online repository[305].

Modeling of irAEs and Recovery

The severity of irAEs is described in grades ranging from grade 0 to grade 5. The absence of an irAE is defined as grade 0, whereas grade 5 is used to represent death caused by an irAE[306]. Within our model, grade 3, 4, and 5 irAEs are aggregated in a "grade3_and_4" irAE, because all of these grades of irAEs manifest with severe physical ailments that requiring clinical management. IrAEs progress from stage 1 to stage 3 within prespecified time intervals. The described time intervals are based on the average time between irAE development and presentation of severe physical symptoms (i.e., grade 3 irAEs). Therefore, patients transition from grade 1 to grade 2 and from grade 2 to grade 3 in 0.5 times the time it takes from irAE development until symptomatic disease (i.e., grade 3 irAEs; Table 2, "Time between Development and Grade 3–4 AE" column).

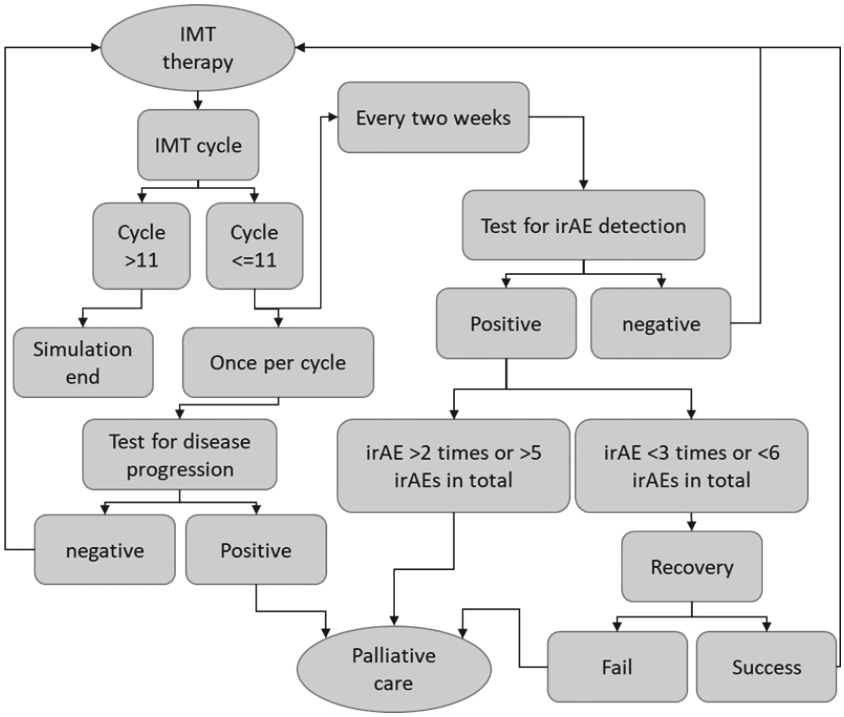


Figure 1 - High-level overview of the clinical pathway. One IMT cycle consists of 6 weeks of treatment with nivolumab. A test to detect progressive disease is performed once in every treatment cycle, and tests to detect irAEs are performed every two weeks. Patients diagnosed with progressive disease, incur a specific irAE a third time, or incur an irAE a sixth time transition to palliative care. Solid lines are used to depict standard transition options, the dashed line represents a conditional transition, i.e., the transition depends on the outcome of a separate process, in this case the detection of progressive disease. IMT-immunotherapy; irAE-immune related adverse event

As shown in Figure 1, patients in whom 1 of the tests results in a positive outcome can either enter a recovery phase or transition to palliative care. During the recovery period, patients are withheld from IMT for a duration of 5 wk. Patients who are diagnosed with an irAE within the first IMT cycle transition to a "fast recovery" state, in which the recovery period is reduced to 2 wk. The fast recovery state is introduced to resemble clinical decision making during the first IMT cycle. During the first cycle, physicians strive to optimize the chances of IMT to have a beneficial effect. During recovery, the test for disease progression continues according to the 6-wk schedule, and patients diagnosed with progressive disease during recovery transition to palliative care directly. Recovery from irAEs in the recovery phase is dependent on a prespecified probability. This recovery probability is based on the IMT cycle number, the grade of irAE, and the number of previous irAEs (Table 3). A transition to palliative care is made when recovery from the irAE fails. IMT treatment is ceased indefinitely after entering palliative care. Moreover, the model allows for recovery of the same type of irAE twice, and patients are allowed to recover 5 times from any combination of irAEs included in the model. In case a specific irAE occurs for the third time or a patient develops an irAE for the sixth time, a transition to palliative care is made directly without entering the recovery phase.

Model Calibration

Because the accuracy of the diagnostic pathway (i.e., the combined accuracy of the tests and interpretation of test results by a physician) aimed at the detection of irAEs is unknown because of a paucity of information regarding the accuracy of tests in this specific application (i.e., the detection of irAEs and the influence of a physician interpreting these test results), model calibration is performed to improve the accuracy of the diagnostic path and ensure the internal validity of the model by comparing the model output to real-world patient data. The model was calibrated by changing the input values for the sensitivity and specificity of the diagnostic pathway and comparing the probability of patients entering palliative care over time. The cumulative probability of patients entering palliative care over time is calculated based on real-world data using R statistical software version 3.6.1 and the ecdf function included in the stats package[307, 308]. Within UPPAAL, a query is run to simulate 11 treatment cycles (i.e., 66 wk) and 100,000 patients. The probability of entering palliative care over time is retrieved directly from the query output in UPPAAL. Model outputs are compared with real-world data through data visualization using the ggplot2 package (version 3.2.1) in R[309]. The model calibration is considered successful when the model outputs are within the confidence bounds surrounding the real-world patient data. Confidence bounds are generated according to the Dvoretzky–Kiefer–Wolfowitz inequality[310].

One of the shortcomings of using only the test outcomes is the incapability of expressing the overall physical state of the patient, which might be of great influence concerning the decision on treatment continuation. In practice, test results are interpreted by a physician; this interpretation step is likely to result in an increase in the accuracy of the diagnostic process. Herein we define the accuracy of the diagnostic process as the accuracy of the test after interpretation of the test results by a physician. The accuracy of the diagnostic process is expressed in terms of test sensitivity and specificity. Conversely, the test accuracy is still used to refer to the sensitivity and specificity of each test when

outcomes are solely compared with the threshold values for disease detection. Since the accuracy of the diagnostic process is unknown, we adjusted this accuracy until the described model outcomes closely match the observed patient data.

Scenario Analysis

A scenario analysis is performed to assess the influence of changes in diagnostic accuracy, that is, the sensitivity and specificity of the diagnostic process on the probability of patients entering palliative care within 66 wk of IMT. The scenario analysis makes use of a query, which provides the probability of patients entering palliative care within 66 wk. For this scenario analysis, 14 scenarios with different input values for the test sensitivity and specificity are drafted, including 2 sensitivity values to represent a high and low test sensitivity. The scenario analysis includes 7 specificity values chosen after empirical tests during model calibration show that a specificity lower than 88% results in a probability of 1 that patients would enter palliative care before week 66.

Results

Model Construction

As described in the Methods section, the constructed model consists of 6 templates, with each template fulfilling a specific function in modeling the clinical pathway. Here, we describe 2 templates in more detail to provide insight into the inner workings of the model.

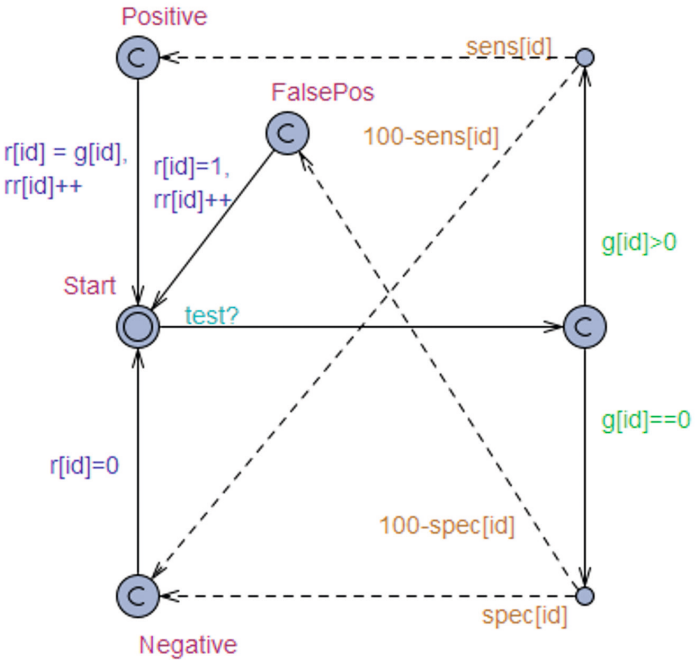


Figure 2 - Template "Test", simulation of tests aimed to detect irAEs based on the test sensitivity, specificity, and presence of an irAEs. Solid line: transition path, dashed line: transition based on a probability of that line being executed. Green text: Guards, a requirement that must be met to allow for the transition to occur. Orange text: Probability, the probability with which a transition will occur. Blue text: update, once the transition occurs the defined parameters will receive an update. Light blue text: synchronization, the transition name followed by a "?" is a receiving channel and the transition will take place once the synchronization signal is received. If the name is followed by a "!" the channel will be used as a broadcasting channel and a synchronization signal will be send once the transition occurs.

Figure 2 depicts the template used to model the 6 different diagnostic processes that correspond to the irAEs included in the model. The template is replicated 6 times during a simulation, and each copy of the template is assigned an irAE through an identifier ([id]). Each test template is initiated in the "Start" location, and all tests are performed simultaneously when a signal is received through the synchronization channel indicated by "test?" When this synchronization signal is received, a transition is made from the start location to the location indicated by "irAE present?" The test outcomes depend on the presence of an irAE; in the model, "g[id]" is used to indicate the grade of irAE for each of

the irAEs included in the model. In case a patient presents with the irAE corresponding to the respective test, the patient will follow the path "g[id]>0." Patients free of the irAE continue through the path indicated by "g[id]==0." For patients who present with an irAE, the test outcomes depend on the test sensitivity defined by "sens[id]." Patients presenting with an irAE transition to the location "Positive" in case of a positive test result. The probability of this true-positive test result is equal to the test sensitivity, whereas the probability of a false-negative test result (i.e., a transition to the location "Negative") is equal to 1 minus the test sensitivity. Patients free of an irAE follow the path downward from "irAE present?" and can receive either a true-negative (transition to the location "Negative") or false-positive test result (i.e., transition to the location "FalsePos"). The probability of receiving a true-negative test result equals the specificity of a test (spec[id]). The probability of receiving a false-positive test result is equal to 1 minus the test specificity. From the location "Positive,""FalsePos," or "Negative," the patient returns to start. This transition automatically updates the values of "r," which represents the grade of irAE incurred, and "rr," which represents the number of times a specific irAE occurred.

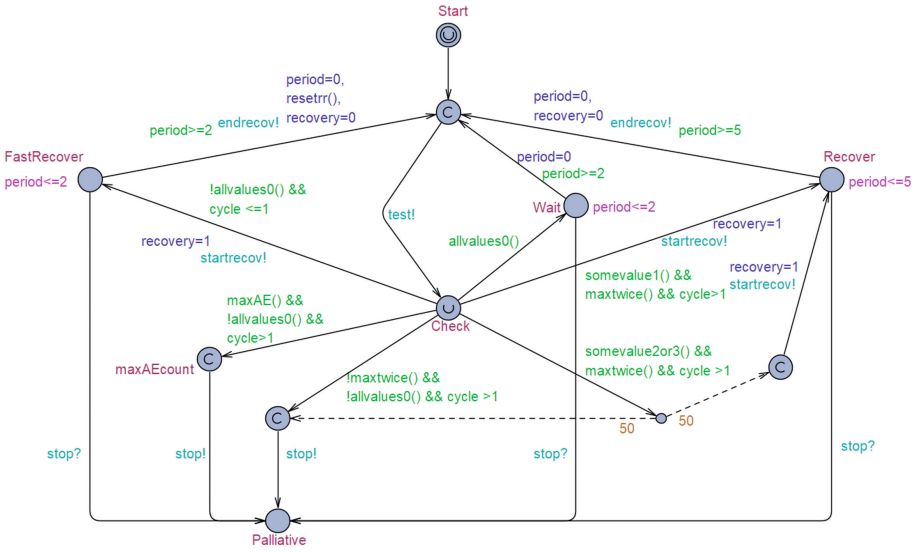


Figure 3 - Template "Protocol", simulates the test protocol and clinical decision making. Solid line: transition path, dashed line: transition based on a probability of that line being executed. Green text: Guards, a requirement that must be met to allow for the transition to occur. Orange text: Probability, the probability with which a transition will occur. Blue text: update, once the transition occurs the defined parameters will receive an update. Light blue text: synchronization, the transition name followed by a "?" is a receiving channel and the transition will take place once the synchronization signal is received. If the name is followed by a "!" the channel will be used as a broadcasting channel and a synchronization signal will be send once the transition occurs. Pink text: invariant, an upper limit for the maximum time until the next transition has to occur from this location.

The test protocol template as depicted in Figure 3 is initiated in the “Start” location. This template is used to simulate the IMT treatment cycles and interpretation of test results. The actions taken in the test protocol depend on a time in weeks indicated by “period.” During IMT, patients receive tests every 2 wks aimed at the detection of irAEs. The tests are performed when the patient transitions from the location “Neutral” to the location “Check.” During this transition, the communication channel “test!” is activated. Test results are evaluated in the location “Check.” Depending on the test result, patients can either continue the standard test sequence when no irAEs are found, that is, the patient transitions to the location “Wait” through the path indicated by “allvalues0()”. Go into a fast recovery phase “FastRecover” if the patient is diagnosed with an irAE during the first treatment cycle (path: !allvalues0 && cycle<=1). Enter the normal recovery phase “Recover” if the patient is diagnosed with a grade 1 irAE, is diagnosed with an irAE fewer than 3 times, and has completed at least 1 IMT cycle (path: somevalue1() && maxtwice() && cycle >1). Patients diagnosed with a grade 2 or 3 irAE, who have been diagnosed with an irAE fewer than 3 times, and who have completed at least 1 IMT cycle have an equal probability of 0.5 of either entering the recovery phase (location: Recover) or transitioning to palliative care (location: Palliative). This probability of 0.5 is indicated by the number 50 near the dashed arrows. Patients who receive a third positive test result for one of the included irAEs or who are diagnosed with an irAE for the sixth time transition to palliative care (location: Palliative) directly. The fast recovery period is defined to last 2 wk, as defined by the guard “period>=2” and the invariant “period<=2”, meaning the transition has to occur when the value of period equals 2. The standard recovery period is defined to last 5 wk (guard: period>=5, invariant: period<=5). The probability of recovery depends on the grade of irAE and the number of times the patient is diagnosed with the irAE. This probability is looked up in a table using the notation c[g[id]][rec[id]], in which “c” indicates the probability of recovery based on “g,” which represents the grade of irAE, and “rec” represents the number of real detected irAEs (i.e., the number of previously incurred true-positive test results).

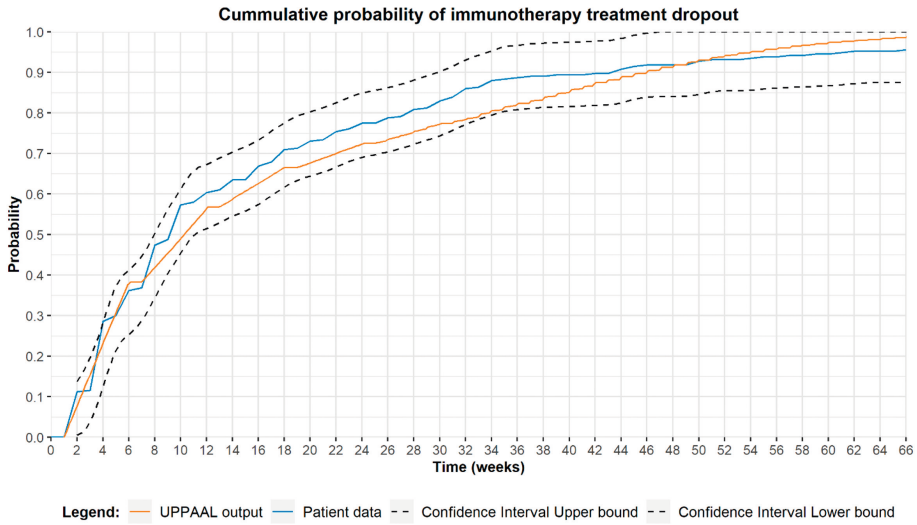


Figure 4 - Model calibration, the probability of patients transitioning to palliative care over time. The blue line represents patient data, the orange line depicts model output, and the black lines represent the confidence bounds surrounding the patient data based on the Dvoretzky–Kiefer–Wolfowitz inequality. The model calibration was performed using the accuracy of the diagnostic process, satisfactory results (i.e. model outputs are located within confidence bounds surrounding the patient data over the full 66 week period) were provided using a sensitivity and specificity of 85% and 91%, respectively. Model outcomes were derived using the query: E[<=66;100 000](max:paltime)

Model Calibration

The model was calibrated by comparing the cumulative probability distribution of patients entering palliative care over time to observed patient data. Ultimately, a sensitivity and specificity of 85% and 91% provided a satisfactory fit, respectively. The choice was based on the visual fit of the model outcome as compared with patient data and its corresponding confidence bounds. However, after calibration, the model still underestimates the probability of patients entering palliative care slightly between week 8 and 48 of IMT and overestimates this probability from week 50 until week 66. Figure 4 depicts the cumulative probability distribution of patients entering palliative care over time, derived from patient data and model outcomes.

Scenario Analysis

A scenario analysis was performed to evaluate the effect of the accuracy of the diagnostic process on the probability of patients entering palliative care. Table 4 depicts the probability of patients entering palliative care within 11 IMT cycles given a combination of sensitivity and specificity values for the diagnostic process (i.e., test results, including interpretation of test results by a physician). Our results show that changes in test specificity can have a significant effect on the probability of patients entering palliative care within 11 IMT cycles, with a difference of 15% between a test specificity of 88% and 99% (Table 4). Moreover, there was no significant difference between the scenario with

the high sensitivity and low sensitivity in patients entering palliative care before week 66 of IMT, ceteris paribus.

Table 4 - Outcomes of the scenario analysis. The probability of patients transitioning to palliative care within 11 IMT cycles given a pre-specified combination of sensitivity and specificity of the diagnostic path. The specified diagnostic accuracy of the diagnostic path was applied to all six tests corresponding to the six immune related adverse events included in the model. Model outcomes were derived using the query Pr[\=66](\.Protocol.Palliative). The top row represents the test specificity, the left most column represents the two scenarios including a high and low sensitivity.

		Specificity						
Problability of patients transitioning to palliative care with 11 treatment cycles.		**.88**	**.90**	**.92**	**.94**	**.96**	**.98**	**.99**
Sensitivity	**.60**	0.99	0.99	0.98	0.95	0.89	0.84	0.83
	.90	0.99	0.99	0.98	0.95	0.89	0.84	0.83

Discussion

Managing IMT-induced irAEs is one of the great challenges in cancer management today. Several attempts have been made to predict patients' susceptibility to irAEs in the treatment of solid tumors using immune checkpoint inhibitors. A variety of biomarkers have been studied in this context, including pretreatment serum antibody levels in melanoma patients[311] and baseline thyroid peroxidase, thyroglobulin, follistatin, and human interferon-inducible protein-10 levels in NSCLC patients [312, 313]. Moreover, a review by von Itzstein et al [314] demonstrated the large variety of biomarkers studied in relation to the diagnosis and prediction of irAEs. Although predictive biomarkers might aid the identification of patients with a high susceptibility for irAEs, monitoring is still needed to identify any occurring irAEs. To our knowledge, there are no previous studies that have aimed to construct a model to analyze the development and detection of IMT-induced adverse events in lung cancer patients. In this article, we present a model that can eventually be used to evaluate the influence of test accuracy, timing, and composition of the diagnostic test panel on treatment continuation. Optimization of the diagnostic test panel might ultimately lead to a cost reduction through less frequent testing or a reduction in biomarkers included in the test panel. In addition, as depicted in Table 1, the current diagnostic panel consists of 30 biomarkers assigned to 1 of 4 analysis categories (i.e., blood count, liver function, clinical chemistry, and special chemistry). Although the current model aggregates all relevant biomarkers and their interpretation into a single test per irAE, an extension of the proposed model could be used to evaluate the added value of each individual biomarker in the diagnostic panel. Currently, a physician needs to analyze results from all biomarkers included in the diagnostic panel, resulting in a complex decision scheme. However, optimization of the diagnostic panel results in removal of tests with little added value from the test panel, resulting in a potentially less complex decision scheme. Although optimization of the test sequence would be unfeasible in prospective or retrospective studies because of the number of possible diagnostic strategies, a modeling approach allows for the evaluation of a large number of test sequences with less financial and time constraints. Our results indicate the feasibility

of developing and calibrating a TA-based model developed in UPPAAL to simulate IMT in lung cancer patients, including disease progression and development of irAEs.

During model development, the probability of completing all 11 IMT treatment cycles was chosen as the outcome parameter to calibrate the model and compare different scenarios. This parameter was chosen in combination with the time horizon, which was limited to the IMT treatment period. It is known that the care pathway in NSCLC is very heterogeneous, and various treatment options are available after IMT. This heterogeneity in therapeutic pathways makes it unfeasible to include all relevant pathways in the model and to incorporate the effect of IMT on quality of life and survival during subsequent treatment lines. In the model, it is assumed that a 5-wk treatment cessation during a recovery period does not influence treatment outcomes. However, ceasing treatment too early might result in withholding a potentially beneficial treatment from patients.

With regard to model calibration, the model still slightly underestimates the probability of patients transitioning to palliative care during the period of week 8 to week 48 of treatment. Conversely, the model slightly overestimates this probability for the remaining 18 wk. These differences might be explained by 3 modeling challenges.

First, the diagnostic process of irAE detection is difficult, and there is a lack of strict guidelines on restarting IMT after a patient recovers from an irAE. Second, irAEs occur with a relatively low incidence, and patients are more likely to stop IMT because of disease progression. This might result in an underestimation of the probability of developing irAEs or the probability of recurrence after recovery. In the data set used during model construction, the percentage of incurred irAEs ranged from 0.4% up to 6.5% for individual irAEs, whereas in total, 18.1% of patients incurred an irAE. This low prevalence directly affects the uncertainty regarding the timing of events and the probability of occurrence. However, this does not affect the viability of UPPAAL in modeling the clinical pathway or the ability to calibrate the model. Third, little is known about the actual sensitivity and specificity of the diagnostic process in this specific setting, mostly because thresholds for disease detection are derived from other patient groups or a healthy population. In addition, because test results contribute to the decision-making process, the sensitivity and specificity of the diagnostic process can never be 100%, as many other factors influence the decision on treatment continuation. Moreover, within this process, it is unclear how much the true diagnostic accuracy of the tests is influenced by the physician, since the data used during model construction provide information only on the actual clinical decision. Although the true accuracy of the test is unknown, the model could be used to assess the influence of removing a test or changing the test frequency on the probability of completing all IMT cycles on cohort level.

We performed the scenario analysis to evaluate the influence of the accuracy of the diagnostic process on the probability of patients transitioning to palliative care. This scenario analysis shows a strong influence of test specificity, although the influence of test sensitivity did not appear to affect model outcomes. These results match our expectations given the low incidence of irAEs. The absence of an effect of test sensitivity might partially be explained by the low incidence of irAEs. However, the relatively

high-test frequency is also likely to lower the effect of changes in test sensitivity on model outcomes, as this could limit the impact of false-negative test results on model outcomes. Moreover, the direct effect of test sensitivity is not fully captured by the outcome measure, because a low-test sensitivity might only delay the time until the irAE is detected in the model. This delayed detection could occur through a positive test later on or due to a symptomatic presentation of the irAE. Hence, the influence of the sensitivity of the diagnostic path on the probability of patients entering palliative care before week 66 is limited. However, there is a difference in the recovery probability depending on the grade of irAE at the time of diagnosis. Therefore, a high-test sensitivity might allow for the earlier identification of irAEs (i.e., lower grade), resulting in more patients successfully recovering from the irAE and completing the 66 wk of IMT. Unfortunately, the low incidence of irAEs reduces the effect size of the sensitivity on cohort level.

In the currently used protocol, patients provide a blood sample every 2 wk. However, the blood samples are obtained at random time points throughout the day, depending on the patient's appointment. The accuracy of the test might be influenced by this inconsistency in timing, since it is known that some blood values fluctuate during the day because of biological variability or under the influence of external factors (e.g., food or beverages). With the introduction of a new test to the test panel, it is key to keep an eye on the influence of the test on clinical decision making. A diagnostic test will provide added value only if it provides actionable results and when a physician acts on these results, in combination with tests already used. In patients with stable disease or response, it might be detrimental to stop IMT because of an irAE in cases of a lower grade or relatively harmless irAE.

Modeling the care pathway and evaluating treatment protocols might be helpful for identifying the most optimal test strategy, based on the composition of the diagnostic kit. Moreover, it would be unfeasible and potentially unethical to evaluate all options in a trial-based setting. Herein we present how UPPAAL can be used to develop a model that emulates the clinical pathway. As with most model-based evaluations, the generalizability of the results strongly depends on the underlying data, the alignment of the model structure with real-world clinical guidelines and pathways, and assumptions about the prior knowledge of physicians using the diagnostic information. Although we do not expect significant differences in the prevalence of irAEs or progressive disease, the clinical pathways and the prior knowledge of physicians do differ between health services. Because the diagnostic accuracy largely depends on the interpretation of results by a physician, and clear guidance regarding the interpretation of the test is lacking, it is expected that there will be differences in management between physicians, not only on an international level but also on an institutional level. In this study, the diagnostic accuracy was estimated in the model calibration and reflects the average accuracy of the diagnostic path for the group of physicians involved in the treatment of the study cohort. It is likely that physicians working in another health service have different experience and prior knowledge, and the generalizability of the current study critically depends on the extent of clinical expertise and variation in prior knowledge.

Future work will involve expanding the model to identify the optimal diagnostic strategy in terms of costs and outcomes.

In conclusion, we have shown that it is worthwhile to construct a TA-based model to emulate complex clinical decisions in the management of NSCLC using UPPAAL. Based on assumptions that can be changed and adapted in the model, we calibrated the model using real-world data. The scenario analysis indicated that the effect of test accuracy on the probability of lung cancer patients treated with immunotherapy transitioning to palliative care is predominantly dependent on the test specificity. Moreover, the influence of test sensitivity is limited, and a high test specificity is important to prevent the too-early termination of IMTs.

General discussion and future perspectives

For this thesis, the framed research question was: "*is it possible to predict a response with one or more factors which directly contribute to immune response, or is an overaching surrogate marker, containing multiple factors, more predictive?".* The upcoming discussion answers this question.

Since immunotherapy is available in clinical practice, the outcome of patients with lung cancer improved in the terms of overall survival and progression free survival, however, it is an expensive treatment and the response rate remains relatively low [3, 5, 42, 133]. In order to maintain a sustainable health care system, further steps have to be taken. Besides Heathcare Technological Assessments, a better selection of patients is of paramount importance. Therefore, biomarker development is important, which allows a better therapy selection for each individual. Allthough the results in this thesis are promosing, the problem is still not solved. Especially in theirfinal stage, patients prefer to retain the best quality of life. With accurate patient selection and monitoring during treatment, it helps to prevent prescribing ineffective treatment.

Until now there are three U.S. Food and Drug Administration (FDA) approved biomarkers for the prediction on response to immunotherapy: PD-L1 (Programmed Death-Ligand 1), tumor mutational burden (TMB) and microsatellite instability high (MSI-H)/deficient mismatch repair (dMMR). The first two FDA approved markers are set out in table 1, where the seven requirements of a perfect marker are compared between the markers. Also, the other potential biomarkers presented and discussed in the previous chapters are expanded. In this thesis, most markers are compared to the current FDA approved marker PD-L1. However, a mutual comparison between all markers would be interesting as well. Table 1 is based both on literature and experience and is intended to guide the comparison, and more importantly, further discussion.

The third FDA approved marker, MSI-H/dMMR, is not included in this comparison. This marker, or a combination of markers, shows a characteritisc of a tumor in which there is a deficiency in the mismatch repair (or dMMR). This leads to cells not being able to recognize and repair spontaneous mutations, leading to a high mutation burden and a high microsatellite instability (or MSI-H) [315]. Since this marker is mostly used in colorectal carcinomas and not (widely) used for prescribing immunotherapy for lung cancer [315, 316], this marker is not included in table 1.

FDA approved markers

PD-L1

The aim of immunotherapy is to active tumor-specific cytotoxic T lymphocytes(CTL), in order to enhance and maintain a good antitumor response [43]. Multiple cells and receptors are involved in this complicated process. Nivolumab and pembrolizumab, both PD-1 checkpoint inhibitors, interact with the PD-L1 receptor, causing such a CTL response [15, 317]. PD-L1 is a biomarker designed for the prediction on response to anti-PD-1 therapy with expressing its target. High expression of PD-L1 in a biopsy of the tumor (primary or metastases), analyzed using immunohistochemistry (IHC), suggests

a good response to immunotherapy [20]. Across different trials evolving nivolumab, objective response (OR) and longer duration of response (DOR) have been registered both in PD-L1-positive and PD-L1-negative NSCLCs, even in numerically higher among positive tumors [2, 36], and no differences have been described for different levels of PD-L1 expression [2, 3, 5, 37]. The predictive role of PD-L1 expression for pembrolizumab showed a limited objective response rate (ORR) of 30%. In treatment naive patients with a tumor PD-L1 expression of >50%, the response rate was 44.8% when treated with pembrolizumab [42]. These positive results resulted in two FDA approvals for registration of pembrolizumab. Besides the limited accuracy of PD-L1, a discordance between two biopsies from the same patient has been observed. This uneven distribution of the PD-L1 positive tumor cells is a known phenomenon in clinical practice [52-57].

TMB

Tumor Mutational Burden (TMB) is defined as a total number of somatic mutations per coding area or per mega Base of a tumor genome [318] and is a potential marker for the prediction of response. In general, smokers show a higher mutational burden than never smokers [115]. The rationale is based on the observation that mostly tumors developed from chronic mutagens (cigarette smoke in lung cancer) show better responses than other tumors [112, 115, 319]. Rizvi et al investigated in two small independent cohort the correlation between tumor mutational burden and response to immunotherapy. They showed that a higher TMB was associated with clinical efficacy of pembrolizumab, with a median of 302 mutations in patients with response, compared to 148 in patients with no response (Mann-Whitney P=0.02)[112]. In the KEYNOTE-158 study, the TMB of patients was evaluated and a high TMB (>10 mutations) was compared to a non-high TMB (<10 mutations). 102 of the 790 evaluable patients showed a high TMB. Here, the overall response rate was 29%. The other 688 patients showed a response of only 6% [318]. Based on this trial, the FDA approved TMB for the treatment of adult (and pediatric) patients with unresectable or metastatic high TMB solid tumors, that progressed following prior treatment and who have no satisfactory alternative treatment options [316]. However, the disadvantages are: it is invasive (a biopsy is needed) and the test is quite expensive compared to PD-L1 [320]. Besides, different trials show a potential improved OS or PFS, but are not yet able to predict response [321].

Table 1 – Simplified overview of the different biomarkers for the prediction or monitoring of immunotherapy response in patients with NSCLC.

	FDA approved	FDA approved	Ch 3	Ch 4	Ch 5+6	Ch 7	Ch 8	Other*
Level of Evidence [322]	1A	2A	2B	2B	2B	3B	2B	2B
Requirement	PD-L1	Tumor mutational burden	Platelet-RNA	Protein classifier	eNose	Dog	Tumor markers	ctDNA
1. Understandable rationale								
2. Accurately predict or monitor a responder								
3. Minimally invasive								
4. Easy to collect and perform								
5. Reproducible, robust and repeatable								
6. Fast								
7. Costs								

*ctDNA is considered the most important other player in the field of non-invasive biomarker research and comparable to the markers in this thesis. The other FDA approved marker, MSI, is not included, since its approval does not regard lung cancer.
** In table 1 in the introduction "cost effectiveness" is considerd an important requirement for a perfect biomarker, however, it is complicated to measure and therefore needs a study on it's own. Therefore, the first derivative, the costs of a marker, is used in this table.
Abbreviations: Ch: Chapter
The traffic light colours indicate requirement available (green), not fuly available (orange) or not available (red).

Biomarkers in this thesis

In **Chapter 3** we investigated whether platelet RNA signatures may provide classification power for nivolumab immunotherapy response prediction before start of treatment. Because platelets are involved in the immune response, we expected that this would allow us to predict whether someone would respond or not. The benefits of this biomarker would be that it is minimally invasive and easy to collect. Therefore, we analyzed and build an algorithm with the blood of 286 patients, all drawn before start of immunotherapy treatment. In a validation cohort of 107 patients, we found an area under the curve (AUC) of 0.58 (95%-confidence interval (CI) 0.45-0.7), which is not sufficient for clinical practice. As discussed before, more samples or another algorithm might improve the use of this potential biomarker. Also, nowadays the result may take days until one week, however, in the nearby future it might be possible that that time is shortenend.

All in all, it is the least promosing marker for the near future for the prediction of response. It would be interesting to know if we would be able to build a classifier for monitoring of response.

The second biomarker investigated is the use of proteins for the development of a classifier. Therefore, in **Chapter 4**, mass spectrometry (MS)-based proteomic analysis was performed on pretreatment sera derived from 289 patients with advanced NSCLC treated with nivolumab. Machine learning combined spectral and clinical data to stratify patients into three groups with good ("sensitive"), intermediate, and poor ("resistant") outcomes. Duration of response and survival were examined, which appeared to differ significantly between the three groups: significantly better OS was demonstrated for "sensitive" relative to "not sensitive" patients treated with nivolumab; HR, 0.58 (95% confidence interval, 0.38–0.87; $P = 0.009$). Our study showed that proteins used for this classifier, were associated with inflammatory processes and wound healing cascades. This is in line with all data reported so far. In this project, multiple validation cohorts where used, containing samples from other hospitals and/or intended treatment group (chemotherapy). This implies an accurately prediction test, which is reproducible, robust, and repeatable. Of note, the test-time of 15 minutes is quite fast and applicable in the clinic.

Chapter 5 and 6 show that exhaled breath analysis by the eNose discriminate between responders and non-responders to anti-PD-1 therapy in NSCLC. Ineffective anti-PD-1 therapy could potentially be prevented in 24% of the patients without erroneously withholding anyone effective treatment. These patients, classified as non-responder might be saved from unnecessary delays and start treatment with a better alternative. In the validation set, results were compared to PD-L1, which showed that the eNose outperformed PD-L1. The rationale is clear: the use of volatile organic compounds (VOCs), originating from metabolic processes represent changes in the immunological reaction. With the use of exhaled breath, it is a non-invasive biomarker. The eNose is designed for its easy use and fast result. Therefore it can be used in clinical practice. Unfortunately, the robustness depends on the sensors of the eNose, which are relativity easy out of balance, for example with temperature changes or the use of alcohol nearby. In an outpatient clinic this would not be a problem. Daily calibration is performed for the check-up of the quality of the eNose. Only the clear rationale is lacking for the eNose as a biomarker, therefore this is a very promising marker for the near future. However, these breath-based technologies have not been introduced in clinical practice yet and there are many hurdles to take prior to implementation.

In **Chapter 7,** the use of a dog to identify medical situtions has been introduced by others. In the article, referred to in this editorial, a trained dog was used for the (early) diagnosis of lung cancer. With their outstanding good scent, dogs are able to smell the small differences between patients with lung cancer and healthy individuals. The results of using a dog in clinical practice are promising, with a sensitivity of 95% and a specificity of 98% for the detection of cancer. It is a minimal-invasive method for the patient since exhaled breath is used. In this study, breath samples are used, which might be a hard to perform, however, using a 'real dog' would make it easy to perform and fast. However, for every dog the full analytic and clinical validation should be performed. Also, how to make sure their results remain stable. What if a dogs get sick and recovers (for example: COVID), should the full validation 'program' start all over again? Therefore, also the costs should be questioned, since living beings need more than just a place to be.

All in all it exemplifies the potential of breath-based diagnostics with, but will never become available due to the completely impracticle nature of this "biomarker".

In **Chapter 8**, readily available tumor markers were used for the monitoring of response at 6 weeks. These markers were introduced decades ago in different laboratories across the world, but so far it has not been possible to use these markers in practice for predicting response. In our study, a relatively easy test was used on two conditions per value (a 50% increase and above a minimal value). In the validation set, a specificity of 91.9% and a sensitivity of 40.2% could be reached with the use of 3 tumor markers. With a relatively easy test based on two conditions per value (a 50% increase and above a minimal value). These results indicate that serum tumor markers can be used to identify patients in which treatment can be discontinued early because of its ineffectiveness. Since the markers are associated with tumor mass and therefore markers are expected to increase when progression occurs, the rationale is clear. It is minimally invasive, easy to collect, already validated for its robustness, and fast. This biomarker cannot be used to predict responses accurately but is a generic monitoring biomarker and as such can be used to stop treatment early on before radiologic or clinical progression and as such can prevent toxicity and costs. The optimal positioning is currently being studies in various prospective trials.

ctDNA as marker

Table 1 also shows another important and comparable player in the field of biomarker development, namely, the use of ctDNA. ctDNA is a byproduct of dying cancer cells and with its short half-life time it can be used for response monitoring of the tumor higher specificity then serum tumor markers [250, 323] . Goldberg et al showed in 28 patients the dynamics of ctDNA in immunotherapy, where they saw a drop of ctDNA in patients who showed a response to immunotherapy [250]. They found a strong agreement between imaging response monitoring and ctDNA response monitoring. The rationale is not specific for immunotherapy; it represents tumor load and is therefore a generic biomarker for response monitoring [324]. Since it can be measured in blood, the test can easily be performed, but it often takes one week or more. ctDNA as a marker is already widely validated. Currently, the test is expensive, and a Heath Technology Assessment should be performed to determine its place in the decision tree. Another disadvantage is a baseline mutation marker must be present for follow up. Its current development and its minimal-invaseness by using blood makes this marker interesting when compared with the other biomarkers. In the nearby future it might be possible to shorten the time to result and the costs. More research is needed to assess the robustness.

Comparison of biomarkers

In search of the perfect biomarker, the candidate markers must outperform the FDA approved markers as shown in table 1. Both approved markers have a clear rationale for the use in IO treatment. However, they are not very accurate, needs invasive testing and some come with high costs (TMB). The PDL1 en TMB biomarkers were developed as companion diagnostics and therefore been implemented early on as standard to allow the use of IO drug therapy [38, 41, 316].

Costs
The costs of the markers vary widely and goes from relatively cheap tumor markers to expensive DNA analysis. Although a cost-effective biomarker is preferred in a sustainable health care system, research is still done with relatively expensive methods. It goes without saying that it is easier for a cheap marker to be costeffective and with that the question of how useful it is to experiment with expensive methods. However, if a marker prevents patients from having severe side-effects, these benefits may still outweigh these costs.

Review

We have tried to address the question if it is possible to predict an IO response to checkpoint inhibitors while using surrogate biomarkers. Therefore two main hypothesis were formulated:
(1) It is possible to predict response with a surrogate biomarker.
(2) There are more perfect biomarkers for the prediction of response to immunotherapy in patients with advanced NSCLC.

This thesis showed that it is possible to predict response with the use of different surrogate markers (table 1), but these are still in the beginning of their development and more research needs to be done.

Strengths
The strength from these beforementioned studies, is based on the high number tested and at the start of treatment. This results in a uniform group and limits certain biases, such as selection bias and temoral bias, as much as possible. Also, all studies contain a training and validation group, confirming the results on a small scale. The results are accessable for other researchers who want to use our research for further projects.

Limitations
At first, all these markers are only tested in one trial, so more research needs to be done. As described in chapter 1, there are multiple steps to be taken before a marker can be used in clinical practice, for example a FDA approval. Since the current treatment landscape of NSCLC is changing, this is a challenging.

These studies are at risk for different forms of bias: sample bias, storage or patient selection bias since only patients living in the Netherlands who were able to visit the NKI (i.e. having the money and the perforamce) could participate. In two studies other centers were involved, which might reduce the chance on bias.

Some patients were included before histologic al samples were analyzed for PD-L1 expression or tested with one of the different available assays. Initially, PD-L1 testing was not required and was often missing, leading to a non-uniformity of the patient group.

Recommendations and future perspective

Biobank
The development period of PD-L1, more or less started in 2012 together with the clinical trials, was relatively short, which shows that combining biomarker and drug development is very effective [15]. However, the development of biomarkers becomes more challenging once the drugs are already available. To safe time, and therefore years of giving to much therapy, it would be recommended that the biobank, collected by pharmaceutical companies during phase 1-3 trial, should become available for other further biomarker development directly following registration of the drugs.

Publication of "negative" results
Most projects are finished within four years. The exception here is the platelet RNA. In today's research world it is all about funding and publications: the more publications, the more successful a research group. Therefore, most projects are finished when a publication has followed. However, there is a strong bias towards publication of positive results [325]. Although more attractive for its readers, it also creates an illusionary world in which either research with a negative outcome is not published, or it is written in a way that there is an outcome looks better than it is. We choose to report the results of a negative biomarker study (Chapter 3). We are convinced of the results and feel that negative studies are of great importance to be presented to the scientific community. This also prevents other researchers to embark on fruitless endeavors.

Compound biomarkers
With more and more biomarkers becoming available, the physician has to make a decision based on multiple results. This is complicated when the biomarkers all point into a different direction. In **Chapter 8**, in which we used multiple readily available tumor markers, we searched for the best combination of these markers. In 10 years' time more studies will show which combinations of markers should be used, and if only one marker needs to show a positive result, or all the markers. Because a large cohort is preferred for the search of a (combination of) biomarker(s), also modeling can be used. In **Chapter 9**, using data from 248 patients in combination with the literature, a model was built that was representative of reality regarding duration of treatment and survival. Then, this model as used to see how often we should run tests to detect side effects. Here we show that with least testing we were also able to detect a side effect on time. Using a model, the first steps could be taken for combining markers.

New biomarker development
In 2018, a few years after immunotherapy was introduced in second line treatment, the base of this thesis, the therapeutic landscape changed. Nowadays, pembrolizumab is used in first line treatment in patients with >50% PD-L1 expression. A combination of chemotherapy and immunotherapy is given in the other situations [50]. Other drug development, the use of already excisiting drugs and/or new combinations of drugs continues [328]. For example, the addition of bevacizumab (an anti-angiogenesis agent) to the current chemotherapy and immunotherapy regime, as seen in the IMpower150 study, showed an overall response rate of 61%, and is therefore quite promising [329].

There is growing evidence that there is a correlation between the microbes in the gut and lung, the so-called microbiome, and cancer development [330]. Studies have shown the influence of the microbiome on response to immunotherapy [331-333], and could therefore be potential biomarkers [45].

This continuous change may complicate the development of new biomarkers in the context patient treatment, but helps us to better understand all processes involved. To address this problem, an overarching standardized operation protocol (SOP), in which paired tumor biopsies blood and microbiome samples are collected, biomarker research can continue.

Collaboration

Last but not least, a well-known recommendation that should not be forgotten is collaboration. In this thesis we started a collaboration with another department (chapter 8) or another hospital (chapter 3, 5, 6), with an available laboratory to develop and validate a biomarker. I do believe that this is an important step for biomarker development and the reason for its success. Therefore, it is recommended to focus on two sorts of collaborations: (1) between professions and (2) between universities/ hospitals to solve these clinically relevant questions.

A Appendices

Nederlandse samenvatting

Englisch summary

Dankwoord

List of publications

Curriculum Vitae

References

Nederlandse samenvatting

Longkanker is de belangrijkste oorzaak van sterfte door kanker, met 14.000 diagnoses en 10.000 doden per jaar in Nederland. Voorheen was chemotherapie de enige vorm van (palliatieve) behandeling, met hierbij een korte overleving van slechts maanden. Sinds de komst van immunotherapie is deze overleving flink toegenomen, met in de eerste resultaten een overleving van zelfs enkele jaren.

Bij immunotherapie wordt het immuunsysteem in het lichaam gereactiveerd om de tumor te herkennen en aan te vallen, met het afnemen van de tumor tot gevolg. Deze vorm van immunotherapie kent een laag percentage bijwerkingen in vergelijking met chemotherapie. Echter, mocht er een bijwerking optreden, dan is deze vaak ernstiger dan bij chemotherapie, waarbij een ontsteking van de darm (colitis), lever (hepatitis) of long (pneumonitis) veelvoorkomend zijn. Ook bestaat de kans dat een ziekte die ontstaat doordat het lichaam zichzelf aanvalt (auto-immuunziekte, zoals bijvoorbeeld reumatoïde artritis), verergerd wordt of op vlamt. De meeste bijwerkingen zijn goed te behandelen middens prednison, een medicijn dat deze immuunreactie remt.

In de eerste twee hoofdstukken worden twee veelgebruikte soorten immunotherapie omschreven. In **hoofdstuk 1** en **hoofdstuk 2** worden nivolumab en pembrolizumab als therapie geïntroduceerd en omschreven in termen van werking, effectiviteit, bijwerkingenprofiel en kwaliteit van leven bij patiënten.

Zoals ook omschreven in **hoofdstuk 1 en 2**, reageert helaas slechts ongeveer 20% op deze vorm van behandeling. Immunotherapie is een dure behandeling, met een paar duizend euro per gift. Goede selectie van patiënten is dus belangrijk voor zowel beperken van ernstige bijwerkingen als kostenbesparing. Ook zouden patiënten nog op tijd een andere therapie die wel mogelijk werkzaam is geboden kunnen worden. Een goede biomarker zou hierbij kunnen helpen. Een biomarker is een biologische marker die te meten is in of rondom een patiënt en die kan aangeven of een aandoening of toestand aanwezig is en/of hoe het beloop ervan is. Bekende voorbeelden hiervan zijn het C-reactief proteïne (CRP) in het diagnosticeren en vervolgen van een infectie, de zwangerschapstest voor de aanwezigheid van een zwangerschap en prostaatspecifiek antigeen (PSA) voor de opsporing van een prostaatcarcinoom. Een biomarker die veel gebruikt wordt is de PD-L1, het aangrijppunt voor de werking van de immunotherapie. Voor deze biomarker is een stukje weefsel van de tumor nodig (biopt), welke niet altijd makkelijk te verkrijgen is. Daarnaast geeft PD-L1 niet altijd goed aan of iemand wel of niet gaat reageren.

In deze thesis ga ik op zoek naar de ideale biomarker om te voorspellen of iemand wel of niet gaat reageren op immunotherapie. Een ideale biomarker heeft een duidelijke rationale, is accuraat, minimaal invasief, makkelijk te verzamelen en te analyseren, is robuust en reproduceerbaar, geeft snel resultaat en is kosteneffectief.

De eerste biomarker die onderzocht is, is het gebruik van bloedplaatjes in bloed. Deze bloedplaatjes bevatten verschillende fragmenten van genetisch materiaal (RNA). Eerder

onderzoek liet zien dat dit RNA gebruikt kan worden voor de zoektocht naar kanker in gezonde individuen. Omdat bloedplaatjes ook betrokken zijn bij een immuunreactie, verwachtten wij ook dat we hiermee konden voorspellen of iemand wel of niet zou reageren. In **hoofdstuk 3** hebben we het bloed geanalyseerd van 286 patiënten, allen afgenomen voor de start van behandeling van immunotherapie. De resultaten laten zien dat het gebruik van bloedplaatjes wel beter is in vergelijking met het random selecteren van patiënten, echter niet voldoende om dit in de klinische praktijk te gebruiken.

De tweede biomarker die onderzocht is, is het maken van een algoritme van verschillende eiwitten in het bloed. Eiwitten worden gemaakt voor en gebruikt door verschillende processen in het lichaam, zoals bijvoorbeeld het CRP. In **hoofdstuk 4** werd hiervoor bloed van 289 patiënten geanalyseerd, waarbij alles afgenomen was voor de start van de behandeling. Door alle samples te laten analyseren door twee verschillende testen, waarbij de eerste test bepaalde welke tweede test er volgde, kon er onderscheid gemaakt worden in drie groepen: reageert wel, reageert mogelijk en reageert niet. Er werd hiervoor gekeken naar de duur van response en van de overleving, welke significant bleek te verschillen tussen de drie groepen. Verdere analyse liet zien dat eiwitten die gebruikt werden voor dit algoritme, geassocieerd zijn met verschillende processen van de ontsteking- en wondgenezingscascade.

De derde biomarker die onderzocht is, is die van het gebruik van al uitgeademde lucht. Deze uitgeademde lucht bestaat voor een zeer klein deel uit vluchtige gassen. Deze gassen komt voort uit verschillende processen uit het lichaam, zoals ontstekingsprocessen of de stofwisseling. Deze gassen zouden dus ook gebruikt kunnen worden voor de voorspelling op een reactie. In ons onderzoek is daarvoor een electronische neus (de eNose) gebruikt. In **hoofdstuk 5** hebben we uitgeademde lucht van 143 patiënten die zouden gaan starten met immunotherapie geanalyseerd. We hebben hierbij gekeken of we een groep patiënten kunnen identificeren die niet gaan reageren, zodat je bij deze patiënten een andere therapie zou starten. Met onze test laten we zien dat we behandeling kunnen besparen in 24% van de populatie. In **hoofdstuk 6** hebben wij patiënten geanalyseerd die zowel voor start als na start van de behandeling geblazen hebben in de eNose. Hierbij blijkt dat het toevoegen van de resultaten na start van de behandeling de resultaten zeer verbeterd, met ook hierbij weinig vals-positieven en vals-negatieven. Dit zou gebruikt kunnen worden voor de follow-up van patiënten die net gestart zijn.

De electronische neus behaalt goede resultaten, echter kan de natuur het beter. In **hoofdstuk 7** wordt omschreven wat er zou gebeuren als een de neus van een hond gebruikt zou worden. Een spaans onderzoek laat zien dat er bij de zoektocht naar wel/geen longkanker er vrijwel geen vals-positieve en vals-negatieve resultaten waren. Nadelen zijn onder andere de moeilijkheid om een hond als biomarker geregistreerd te krijgen en de stabiliteit van resultaten bij honden in de loop van de tijd.

De vierde biomarker in deze thesis is het gebruik van tumor markers. Deze markers zijn al jaren geleden geïntroduceerd in verschillende laboratoria, echter is het tot nu toe niet gelukt om deze markers te gebruiken in de praktijk voor het voorspellen van respons. In **hoofdstuk 8** hebben we 376 patiënten geanalyseerd die immunotherapie kregen,

waarbij er vijf verschillende markers geanalyseerd werden (NSE, Cyfra21.1, CEA, CA125 en SCC). Er werd gekeken of we op basis van een relatief makkelijke test op basis van twee voorwaarden (een verhoging van 50% vanaf een bepaalde hoogte van de marker) een voorspelling konden doen of iemand niet zou reageren, bij voorkeur met het gebruik van meerdere markers. Dit bleek redelijk accuraat te kunnen, met het gebruik van Cyfra21.1, CEA met/zonder NSE. Het voordeel van deze test is dat deze al in veel laboratoria gebruikt wordt, goedkoop is en redelijk makkelijk te meten. Het nadeel is dat het aantal vals-negatieven, waarbij de test positief wordt beschouwd op het moment dat iemand niet reageert, nog steeds vrij hoog is. Daarnaast is deze test pas bruikbaar als iemand al gestart is met therapie, wat het minder kostenbesparend maakt.

Samengevat zijn er vier biomarkers onderzocht voor de voorspelling op respons, waarbij er een paar veelbelovend zijn. Naast deze vier zijn er nog vele andere biomarkers in ontwikkeling, naast degene die al in de praktijk gebruikt worden. Het is met deze grote hoeveelheid data dan soms lastig om te kijken welke biomarker je wanneer gaat gebruiken, op welk moment en in welke combinatie. Om dit te kunnen bestuderen, zonder dat hier echte patiënten bij betrokken zijn, is het gebruik van een model. Op deze manier kan je door middel van simulaties duizenden patiënten analyseren met verschillende tests en sensitiviteit/specifiteit. In **hoofdstuk 9** laten we een voorbeeld zien van zo'n model. Met behulp van data van 248 patiënten in combinatie met de literatuur werd een model gebouwd wat representatief was aan de werkelijkheid met betrekking tot duur van behandeling en overleving. Dit model werd vervolgens gebruikt om te kijken hoe vaak we tests moeten doen om te kijken of iemand een bijwerking ontwikkelt van immunotherapie. Hierbij laten we zien dat we met minder vaak testen vaak nog steeds op tijd een bijwerking vangen.

In de discussie worden de verschillende markers nogmaals besproken en met elkaar vergeleken.

Een blik in de toekomst laat zien dat de meeste markers verder onderzocht en doorontwikkeld (moeten) worden. Mogelijk zal het combineren van markers tot betere resultaten leiden.

English summary

Lung cancer is the leading cause of cancer death in the Netherlands, with 14.000 diagnoses and over 10.000 deaths per year in 2019. For years chemotherapy was the only (palliative) treatment, with a short survival of only months. Since the introduction of immunotherapy, this survival has increased significantly, with the first results showing a survival of even a few years.

Immunotherapy reactivates the immune system to recognize and attack the tumor, leading to tumor shrinkage. This form of immunotherapy has a low percentage of side effects compared to chemotherapy. However, if a side effect occurs, it is often more severe. Inflammation of the gut (colitis), liver (hepatitis) or lung (pneumonitis) are common. In addition, when a patient suffers from an autoimmune disease, such as rheumatoid arthritis, there is a chance of a flare of this disease. Most side effects are treatable with corticosteroids, a drug that inhibits this auto immune response.

In the first two chapters, two frequently used immunotherapies are described. In **chapter 1** and **chapter 2,** nivolumab and pembrolizumab are introduced and described in terms of mechanism of action, efficacy, toxicity and quality of life of patients.

As described in **chapter 1 and 2**, only 20% shows a response in this form of therapy. Immunotherapy is expensive, with a few thousand euros per cure. A good selection of patients is important for limiting the development of severe side effect and cost reduction. Also, patients could be offered another therapy in time that may be effective. A biomarker could help. A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic or pathogenic processes, or the response to a therapeutic intervention, as described by Lee et al [26]. Examples are C-reactive protein (CRP) in the diagnostics of an infection, a pregnancy test for the presence of a pregnancy and prostate specific antigen (PSA) for the detection of prostate cancer. A frequently used biomarker for the prediction of response of immunotherapy is PD-L1, the target for immunotherapy. For this biomarker a biopsy is needed, which is often hard to acquire. Besides, PD-L1 is not very accurate.

This thesis is about the search for a new biomarker. Criteria for a perfect biomarker are an understandable rationale, accurately predicts or monitors a responder before treatment, is minimally invasive, is easy to collect and perform, is reproducible, robust, repeatable, fast and cost-effective.

The first investigated biomarker is the use of platelets in blood. These platelets contain different fragments of genetic material (RNA). Earlier research showed that RNA could be used for detecting lung cancer. Because platelets are involved in the immune response, we expected that this would allow us to predict whether someone would respond or not. In **chapter 3** we analyzed the blood of 286 patients, all drawn before start of immunotherapy treatment. Our results show that the used of platelets is a little bit better than to select patients randomly, however, this is not sufficient for clinical practice.

The second investigated biomarker is the use of proteins for the development of a classifier. Proteins are made and used for different processes in the human body, such as CRP in inflammation processes. In **chapter 4** blood of 289 patients were analyzed, all samples were taken before start of treatment. All blood samples were analyzed with two different tests; the result of the first test determined the second test. This resulted in three groups: A group that would respond, that probably would respond and would not respond. Duration of response and survival were examined, which appeared to differ significantly between the three groups. Further analysis showed that proteins used for this classifier were associated with different processes of inflammation and wound healing cascades.

The third investigated biomarker is the use of exhaled breath. Exhaled breath contains for a small part of volatile organic compounds. These gasses arise from various processes in the body, such as inflammatory processes or the metabolism. These gases could therefore also be used to predict response. In our research, an electronic nose (eNose) was used for this purpose. In **chapter 5** exhaled breath from 143 patients were analyzed, all samples taken before start of treatment. Here we hypothesized that we could identify a group that would not respond, so that these patients could start another treatment. With this test we show that we could save treatment in 24% of the patients. In **chapter 6**, data of patients who were sampled both before and during treatment were added. Here we show that adding the results after the start of treatment improves the results, with only a few false-negative and false-positive results. This could be used for the follow-up of patients who just started immunotherapy.

The electronic nose shows great results, however, nature beats that. In **chapter 7** is described what would happen if a dog is used as a biomarker. A Spanish research group shows that, with the use of a dog in the detecting of lung cancer, there was a low number of false-positive and false-negative results. However, disadvantages include the difficulty of getting a dog registered as a biomarker and the stability of results in dogs over time.

The fourth biomarker investigated is the use of tumor markers. These markers were introduces years ago in different laboratories cross the world, but so far it has not been possible to use these markers in practice for predicting response. In **chapter 8**, tumor markers of 376 patients who received immunotherapy were used. Five different markers were used, namely NSE, Cyfra21.1, CEA, CA125 and SCC. Our aim was whether we could predict if someone would not respond based on a relatively easy tests based on two conditions (a 50% increase and above a minimal value), preferably using multiple biomarkers. This turned out to be quite accurate, using cyfra21.1, CEA with/without NSE. The advantage of this test is that it is already available in different laboratories, it cheap and quit easy to measure. The downside it that the rate of false-negatives, with a positive test defined as some who would not respond, is still quite high. In addition, this test is only useful for the follow-up of patients, which makes it less cost-effective.

In summary, there are four biomarkers investigated for the prediction on response to immunotherapy, some of which are promising. Next to these four biomarkers, many other biomarkers are in development, beside the once who are already used in clinical

practice. With this large amount of data, it is sometimes difficult to see which biomarker you will use when, at what time and in what combination. To be able to study this without involving real patients, it could be useful to use a model. This way you can analyze thousands of patients with different test combinations and sensitivities and specificities through simulations. In **chapter 9** an example of a model I shown. Using data from 248 patients in combination with the literature, a model was built that was representive of reality with regard to duration of treatment and survival. Then, this model as used to see how often we should run tests to detect side effects. Here we show that with least testing we were also able to detect a side effect on time.

In the discussion the different markers are discussed and compared. Future perspective shows us that most markers are further developed. Maybe combinding tumor markers will lead to better results.

Dankwoord

"Wat ik vroeger niet kon, wat ik nu dus inmiddels wel kan", de uitspraak van mijn moeder Lia Charité († 01-03-2018) op haar sterfbed. Als je eenmaal door hebt waar het leven echt om draait, een inzicht dat je soms kan krijgen als je de dood in de ogen kijkt bij een kankerdiagnose, is het soms helaas te laat. Met dit proefschrift hoop ik dat degenen met ditzelfde inzicht de kans krijgen er nog iets mee te doen, door ze meer tijd en/of een goede, behouden kwaliteit van leven te geven. Dit proefschrift draag ik dan ook op aan mijn moeder, doordat ik door haar de motivatie voel om onderzoek te blijven doen: er valt nog zoveel te ontdekken. Mam, ik weet dat je er bent: bedankt dat je er was en voor al je steun.

Mijn PhD tijd was een leuke tijd waarin ik veel heb geleerd. Ik ben blij met het resultaat, waarvoor ik graag ook een paar mensen wil bedanken.

Allereerst de patiënten en familieleden voor hun bijdrage in het onderzoek.

Dan uiteraard Michel & Paul.

Michel, dank voor het met mij doorlopen van het traject "van simpele dokter naar onderzoeker". Dank dat je me vrijliet, maar net niet te veel. Wat heb ik daardoor veel geleerd.

Paul, dank voor de vele serieuze longkanker en onderzoeksgerelateerde dingen die ik heb geleerd, maar vooral ook van alle verhalen daaromheen.

En dan degene met wie ik veel samengewerkt heb:
Vincent, dank voor je creatieve inzicht in de statistiek. Sjaak, dank voor al je vertrouwen rondom en naast mijn onderzoek. Daan, dank voor je heldere kijk op de projecten. Egbert, dank voor je ongezouten mening. Huub, dank voor de goede samenwerking. Kim, bedankt voor het inzicht dat pathologie lang niet zo saai is als ik dacht. Ruben, dank voor alle koffie.

De afdeling thoraxoncologie: Bedankt Robert, Karlijn, Irene, Dianne, Merel en Willemijn. Daarnaast dank aan de AIOS, Fellows, longartsen en natuurlijk de VSers.

Giulia, thank you for helping me started with my PhD!

In dit proefschrift staan veel projecten met verschillende onderzoeksgroepen:

Voor het project met de bloedplaatjes wil ik graag uit het AUMC Myron, Nik, Adrienne, Tom, Anna-Larissa, Sjors, Tineke, Lisanne, Pepijn, Jip en Edward bedanken. Leuk dat de publicatie toch nog gelukt is.

Voor het project met de serumeiwitmarkers wil ik graag Joanna & Heinrich Roder, Anna-Larissa en Egbert bedanken.

Voor het project met de electronische neus, wil ik graag uit het AUMC Rianne, Peter, Paul en Yennece bedanken. Daarnaast Alessandra voor het vastbijten in de eNose projecten. In het AvL wil ik graag de longfunctieassistenten bedanken, met in het bijzonder Marguerite.

Voor het project met de tumor markers, wil ik graag Huub, Ruben, Roland, Mirthe, Huub en Daan bedanken.

Voor het project voor de modelering, wil ik graag Rom, Maarten, Veerle en natuurlijk Freek bedanken voor de gezellige brainstorms en het anders/technisch leren denken.

Het trial bureau met Marianne en Marrit en uiteraard Inge Kemp, bedankt. Daarnaast natuurlijk het secretariaat: zonder een goed secretariaat wordt een afdeling niet gerund, dus dank aan o.a. Bea en Bernadette.

Lieve Linde, Martje, Renee, Ilse, Dianne, Joyce, lieve collega's van het AKL, uit het Catharina en uit het Dijklander (Sabrina, Maartje en Emma): De laatste loodjes zijn het zwaarst, maar door jullie zijn ze te dragen, dankjulliewel!

Dan natuurlijk mijn paranimfen. Lieve Irene, dagen ben je getuige geweest van mijn harde werken, maar ook alles daaromheen. Gezellig en heerlijk, maar nu weer tijd voor het serieuze deel. En Sherilyn, als we had kunnen promoveren op het evidence based moeder zijn, hadden we dat zeker cum laude gedaan. Dank dat je tot die tijd in ieder geval mijn paranimf wil zijn, close enough. Dank voor alle steun en interesse.

Dan mijn lieve familie, die in de laatste jaren ook veel heeft meegemaakt: Lieve Saar, Lies & Ria: dank dat jullie er voor mij/ons zijn. Jammer dat Wammes en Kees het niet meer mee kunnen maken.

Veel dank voor mijn vader en broertje, voor alle steun, het oppassen en de heerlijke maaltijden die voor me werden gekookt.

Dan de twee meest belangrijke mensen in mijn leven:

Lieve Myron, nooit gedacht dat ons onderzoek en je praatje op de ESMO ons hier zou leiden. Dank voor al je geduld, hulp, inzicht en inspanningen. Jij bent mijn grote liefde én mijn voorbeeld. Wie weet wat de toekomst ons nog verder brengt. En dan ons positieve resultaat van onze negatieve studie: Lieve Lina, dankjewel voor de motivatie om het af te maken (plus dat je zo'n lieve en snel doorslapende baby bent). Elke minuut zonder jou moet nuttig zijn. Ik hou van jullie.

List of publications

This thesis

Zago G*, **Muller M***, van den Heuvel M, Baas P. New targeted treatments for non-small-cell lung cancer - role of nivolumab. Biologics. 2016;10:103-17.

Muller M*, Schouten RD*, De Gooijer CJ, Baas P. Pembrolizumab for the treatment of non-small cell lung cancer. Expert Rev Anticancer Ther. 2017;17(5):399-409.

Muller M*, Best MG*, van de Noort V, Hiltermann TJN, Nijmeijer ALN, Post E et al. Blood platelet RNA profiles do not enable for nivolumab response prediction at baseline in patients with non-small-cell lung cancer. Currently submitted to tumor biology

Muller M*, Hummelink K*, Hurkmans DP, Niemeijer AN, Monkhorst K, Roder J, et al. A Serum Protein Classifier Identifying Patients with Advanced Non-Small Cell Lung Cancer Who Derive Clinical Benefit from Treatment with Immune Checkpoint Inhibitors. Clinical cancer research : an official journal of the American Association for Cancer Research. 2020;26(19):5188-97.

de Vries R*, **Muller M***, van der Noort V, Theelen W, Schouten RD, Hummelink K, et al. 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 : official journal of the European Society for Medical Oncology. 2019;30(10):1660-6.

Buma AIG*, **Muller M***, de Vries R, Sterk PJ, van der Noort V, Wolf-Lansdorf M, et al. eNose analysis for early immunotherapy response monitoring in non-small cell lung cancer. Lung Cancer. 2021;160:36-43.

Muller M, Baas P. I love my dog. Lung Cancer. 2019;135:228-9.

Muller M, Hoogendoorn R, Moritz RJG, van der Noort V, Lanfermeijer M, Korse CM, et al. Validation of a clinical blood-based decision aid to guide immunotherapy treatment in patients with non-small cell lung cancer. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2021;43(1):115-27.

van Delft F*, **Muller M***, Langerak R, Koffijberg H, Retel V, van den Broek D, et al. Modeling Diagnostic Strategies to Manage Toxic Adverse Events following Cancer Immunotherapy. Med Decis Making. 2021;41(6):693-705.

*Contributed equally

Other publications

Korse CM, **Muller M**, Taal BG. Discontinuation of proton pump inhibitors during assessment of chromogranin A levels in patients with neuroendocrine tumours. British journal of cancer. 2011;105(8):1173-5.

Best MG, Sol N, In 't Veld S, Vancura A, **Muller M**, Niemeijer AN, et al. Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets. Cancer cell. 2017;32(2):238-52 e9.

Schouten RD, **Muller M**, de Gooijer CJ, Baas P, van den Heuvel M. Real life experience with nivolumab for the treatment of non-small cell lung carcinoma: Data from the expanded access program and routine clinical care in a tertiary cancer centre—The Netherlands Cancer Institute. Lung Cancer. 2018;126:210-6.

Wijn DH, Groeneveld GH, Vollaard AM, **Muller M**, Wallinga J, Gelderblom H, et al. Influenza vaccination in patients with lung cancer receiving anti-programmed death receptor 1 immunotherapy does not induce immune-related adverse events. European journal of cancer (Oxford, England : 1990). 2018;104:182-7.

Moritz R, **Muller M**, Korse CM, van den Broek D, Baas P, van den Noort V, et al. Diagnostic validation and interpretation of longitudinal circulating biomarkers using a biomarker response characteristic plot. Clinica chimica acta; international journal of clinical chemistry. 2018;487:6-14.

Hummelink K, **Muller M**, Linders TC, van der Noort V, Nederlof PM, Baas P, et al. Cell-free DNA in the supernatant of pleural effusion can be used to detect driver and resistance mutations, and can guide tyrosine kinase inhibitor treatment decisions. ERJ Open Res. 2019;5(1).

Schouten RD, Egberink L, **Muller M**, De Gooijer CJ, van Werkhoven E, van den Heuvel MM, et al. Nivolumab in pre-treated advanced non-small cell lung cancer: long term follow up data from the Dutch expanded access program and routine clinical care. Transl Lung Cancer Res. 2020;9(5):1736-48.

Trebeschi S, Bodalal Z, Boellaard TN, Tareco Bucho TM, Drago SG, Kurilova I, (...) **Muller M.** (...) et al. Prognostic Value of Deep Learning-Mediated Treatment Monitoring in Lung Cancer Patients Receiving Immunotherapy. Front Oncol. 2021;11:609054.

Aman J, Duijvelaar E, Botros L, Kianzad A, Schippers JR, Smeele PJ, (...) **Muller M.** (...)et al. Imatinib in patients with severe COVID-19: a randomised, double-blind, placebo-controlled, clinical trial. Lancet Respir Med. 2021;9(9):957-68.

In 't Veld S, Arkani M, Post E, Antunes-Ferreira M, D'Ambrosi S, Vessies DCL, (...) **Muller M.** (...) et al. Detection and localization of early- and late-stage cancers using platelet RNA. Cancer cell. 2022.

Hummelink K, van der Noort V, **Muller M**, Schouten RD, Lalezari F, Peters D, et al. PD-1T TILs as a Predictive Biomarker for Clinical Benefit to PD-1 Blockade in Patients with Advanced NSCLC. Clinical cancer research : an official journal of the American Association for Cancer Research. 2022;28(22):4893-906.

van Delft FA, Schuurbiers M, **Muller M**, Burgers SA, van Rossum HH, MJ IJ, et al. Modeling strategies to analyse longitudinal biomarker data: An illustration on predicting immunotherapy non-response in non-small cell lung cancer. Heliyon. 2022;8(10):e10932.

Curriculum Vitae

Mirte Muller werd in 1988 in het Catharina Ziekenhuis te Eindhoven geboren. Haar vwo-diploma, met als profiel natuur en gezondheid en vrije keuze economie 1,2, haalde ze in 2006 op het Luzac college te Eindhoven. Ze werd dat jaar gelijk ingeloot voor de studie geneeskunde, bij haar eerste keuze de AMC-UvA Amsterdam. Na 4 jaar theorie sloot ze haar doctoraal af met een onderzoeksstage in het Antoni van Leeuwenhoek ziekenhuis, onder leiding van dr. B. Taal en Dr. T. Korse, met een onderzoek naar chromogranine A en protonpomp remmers. Hierna nam ze een jaar vrij om te werken en te reizen naar de verenigde staten en Maleisië. In dat jaar bleef ze werken in het Antoni van Leeuwenhoek ziekenhuis op de afdeling bloedafname en radiologie. In juli 2011 startte ze haar co-schappen en legde deze met goed gevolg af. Zij koos voor een keuze co-schap bij de kinderoncologie en sloot af met een stage interne oncologie, deze laatste opnieuw in het Antoni van Leeuwenhoek ziekenhuis. In december 2013 sloot ze het artsexamen met goed gevolg af.

Haar eerste baan was in het Elkerliek ziekenhuis te Helmond, waar ze voor alle beschouwende specialismes (Interne geneeskunde, cardiologie, longgeneeskunde, neurologie en geriatrie) dienst had. Haar eerste baan combineerde ze met een bestuursfunctie, activiteitencommissaris, van de Amsterdamse Studenten Zwemvereniging SPONS (ASZV SPONS). Na een klein jaar startte ze in januari 2015 op de afdeling van Interne Oncologie van het voormalige Daniel den Hoed kliniek te Rotterdam.

In Oktober 2015 begon ze haar promotietraject in het Antoni van Leenwenhoek ziekenhuis, afdeling thorax oncologie, onder leiding van prof. Dr. P. Baas en dr. M. M. van den Heuvel, inmiddels ook professor. Na 4 jaar fulltime onderzoek begon zij weer in de kliniek. Na een korte zoektocht naar een opleidingsplek, waarbij ze onder andere op beide AUMC-locaties heeft gewerkt, begon zij in september 2020 bij de afdeling longgeneeskunde van het Catharina Ziekenhuis te Eindhoven. Hier werd zij in februari 2021 aangenomen voor de opleiding tot longarts. In augustus 2021 startte ze de vooropleiding in het Dijklander ziekenhuis te Hoorn, waarna ze nu de vervolgopleiding tot longarts in het Catharina Ziekenhuis volgt.

Zij woont samen met haar vriend Myron en dochter Lina in Best. In juli 2024 verwachten zij hun tweede kindje.

In haar vrije tijd zwemt en sport zij graag. Daarnaast zet ze zich graag in voor de stichting “Het vergeten kind”.

References

1. 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, 2010. **28**(19): p. 3167-75.
2. Rizvi, N.A., et al., *Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial.* Lancet Oncol, 2015. **16**(3): p. 257-65.
3. Brahmer, J., et al., *Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer.* N Engl J Med, 2015. **373**(2): p. 123-35.
4. Postow, M.A., et al., *Nivolumab and ipilimumab versus ipilimumab in untreated melanoma.* N Engl J Med, 2015. **372**(21): p. 2006-17.
5. Borghaei, H., et al., *Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer.* N Engl J Med, 2015. **373**(17): p. 1627-39.
6. Robert, C., et al., *Nivolumab in previously untreated melanoma without BRAF mutation.* N Engl J Med, 2015. **372**(4): p. 320-30.
7. Topalian, S.L., et al., *Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.* N Engl J Med, 2012. **366**(26): p. 2443-54.
8. Larkin, J., et al., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma.* N Engl J Med, 2015. **373**(1): p. 23-34.
9. Gettinger, S.N., et al., *Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer.* J Clin Oncol, 2015. **33**(18): p. 2004-12.
10. Wang, C., et al., *In vitro characterization of the anti-PD-1 antibody nivolumab, BMS-936558, and in vivo toxicology in non-human primates.* Cancer Immunol Res, 2014. **2**(9): p. 846-56.
11. *IKNL Incidentie Longkanker*. 2019, IKNL: https://iknl.nl/kankersoorten/longkanker/registratie/incidentie, page seen at 9th of september 2022.
12. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
13. Novello, S., et al., *Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2016. **27**(suppl 5): p. v1-v27.
14. Reck, M., et al., *Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2014. **25 Suppl 3**: p. iii27-39.
15. Zago, G., et al., *New targeted treatments for non-small-cell lung cancer - role of nivolumab.* Biologics, 2016. **10**: p. 103-17.
16. Hsu, W.H., et al., *Overview of current systemic management of EGFR-mutant NSCLC.* Ann Oncol, 2018. **29**(suppl_1): p. i3-i9.
17. Kerr, K.M., et al., *Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer.* Ann Oncol, 2014. **25**(9): p. 1681-1690.
18. Pardoll, D.M., *Immunology beats cancer: a blueprint for successful translation.* Nat Immunol, 2012. **13**(12): p. 1129-32.
19. Iwai, Y., et al., *Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12293-7.
20. Doroshow, D.B., et al., *PD-L1 as a biomarker of response to immune-checkpoint inhibitors.* Nat Rev Clin Oncol, 2021. **18**(6): p. 345-362.
21. FDA, A.U., *Expands Approved Use of Opdivo to Treat Lung Cancer*. 2015, FDA: FDA website.
22. EM, A., *New Treatment Option for Patients with Advanced Lung Cancer*. 2015, EMA.
23. Uyl-de Groot, C.A.M.J., R. Zaim, *Nivolumab (Opdivo) bij Gevorderd Plaveiselcelcarcinoom van de long*, I.b.m. gezondheidszorg, Editor. 2015: Instituut beleid & managment gezondheidszorg.
24. Schouten, R.D., et al., *Real life experience with nivolumab for the treatment of non-small cell lung carcinoma: Data from the expanded access program and routine clinical care in a tertiary cancer centre—The Netherlands Cancer Institute.* Lung Cancer, 2018. **126**: p. 210-216.
25. Schouten, R.D., et al., *Nivolumab in pre-treated advanced non-small cell lung cancer: long term follow up data from the Dutch expanded access program and routine clinical care.* Transl Lung Cancer Res, 2020. **9**(5): p. 1736-1748.

26. Lee, J.W., et al., *Fit-for-purpose method development and validation for successful biomarker measurement.* Pharm Res, 2006. **23**(2): p. 312-28.
27. Dobbin, K.K., et al., *Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations.* J Immunother Cancer, 2016. **4**: p. 77.
28. Masucci, G.V., et al., *Validation of biomarkers to predict response to immunotherapy in cancer: Volume I - pre-analytical and analytical validation.* J Immunother Cancer, 2016. **4**: p. 76.
29. Dictronary.com, *Dictronary.com*, in *Dictronary.com*. 2021: https://www.dictionary.com/e/about/; page seen at 23th of july.
30. Jennings, L., et al., *Recommended principles and practices for validating clinical molecular pathology tests.* Arch Pathol Lab Med, 2009. **133**(5): p. 743-55.
31. Ignatiadis, M., G.W. Sledge, and S.S. Jeffrey, *Liquid biopsy enters the clinic - implementation issues and future challenges.* Nat Rev Clin Oncol, 2021. **18**(5): p. 297-312.
32. Simon, R., *Roadmap for developing and validating therapeutically relevant genomic classifiers*. J Clin Oncol, 2005. **23**(29): p. 7332-41.
33. Scheerens, H., et al., *Current Status of Companion and Complementary Diagnostics: Strategic Considerations for Development and Launch.* Clin Transl Sci, 2017. **10**(2): p. 84-92.
34. FDA, A.U. *Companion Diagnostics - FDA*. 2018 12/07/2018 [cited 2022 9/23].
35. FDA, A.U. *Premarker approval (PMA) - FDA*. 2019 05/16/2019 [cited 2022 9/23].
36. Gettinger, S.N., Hellmann, M.D., Spepherd, F.A. et al. , *First-line monotherapy with nivolumab (NIVO; anti-programmed death-1 [PD-1]) in advanced non-small cell lung cancer (NSCLC): safety, efficacy and correlation of outcomes with PD-1 ligand (PD-L1) expression.* . Journal of Clinical Oncology, 2015. **33**(Abstract 8025).
37. Horn, L.R.N.A.M.J., et al. , *Longer-term follow-up of a phase 2 study (CheckMate 063) of nivolumab in patients with advanced, refractory squamous non-small cell lung cancer.*, in *16th World Conference on Lung Cancer (WCLC); September 6–9, 2015*. 2015: Denver.
38. Novotny, J.F., Jr., et al., *Establishing a complementary diagnostic for anti-PD-1 immune checkpoint inhibitor therapy.* Ann Oncol, 2016. **27**(10): p. 1966-9.
39. Herbst, R.S., et al., *Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial.* Lancet, 2015.
40. Garon, E.B., et al., *Pembrolizumab for the treatment of non-small-cell lung cancer.* N Engl J Med, 2015. **372**(21): p. 2018-28.
41. Sul, J., et al., *FDA Approval Summary: Pembrolizumab for the Treatment of Patients With Metastatic Non-Small Cell Lung Cancer Whose Tumors Express Programmed Death-Ligand 1.* Oncologist, 2016. **21**(5): p. 643-50.
42. Reck, M., et al., *Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer.* N Engl J Med, 2016. **375**(19): p. 1823-1833.
43. Borst, J., et al., *Mechanism of action of PD-1 receptor/ligand targeted cancer immunotherapy.* Eur J Immunol, 2021. **51**(8): p. 1911-1920.
44. 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): p. 950-960.
45. Punekar, S.R., et al., *Immunotherapy in non-small cell lung cancer: Past, present, and future directions.* Front Oncol, 2022. **12**: p. 877594.
46. Schalper, K.A., et al., *Objective measurement and clinical significance of TILs in non-small cell lung cancer.* J Natl Cancer Inst, 2015. **107**(3).
47. Tokito, T., et al., *Predictive relevance of PD-L1 expression combined with CD8+ TIL density in stage III non-small cell lung cancer patients receiving concurrent chemoradiotherapy.* Eur J Cancer, 2016. **55**: p. 7-14.
48. Donnem, T., et al., *Stromal CD8+ T-cell Density-A Promising Supplement to TNM Staging in Non-Small Cell Lung Cancer.* Clin Cancer Res, 2015. **21**(11): p. 2635-43.
49. Hummelink, K., et al., *PD-1T TILs as a Predictive Biomarker for Clinical Benefit to PD-1 Blockade in Patients with Advanced NSCLC.* Clin Cancer Res, 2022. **28**(22): p. 4893-4906.
50. NVALT, *Richtlijn niet-kleincellig longcarcinoom.* 2020, Nederlandse vereniging van Artsen voor Longziekten en Tuberculose: https://richtlijnendatabase.nl/richtlijn/niet_kleincellig_longcarcinoom/startpagina_-_niet-kleincellig_longcarcinoom.html.
51. MC, E., *Tarieven ODV per 1 maart 2021*. 2021: https://www.erasmusmc.nl/-/media/erasmusmc/pdf/2-themaoverstijgend/tarieven-odv-per-010321.pdf.

52. Hwang, D.M., et al., *Prevalence and Heterogeneity of PD-L1 Expression by 22C3 Assay in Routine Population-Based and Reflexive Clinical Testing in Lung Cancer.* J Thorac Oncol, 2021. **16**(9): p. 1490-1500.
53. Zito Marino, F., et al., *Heterogeneity of PD-L1 Expression in Lung Mixed Adenocarcinomas and Adenosquamous Carcinomas.* Am J Surg Pathol, 2020. **44**(3): p. 378-386.
54. Naso, J.R., et al., *Discordance in PD-L1 scores on repeat testing of non-small cell lung carcinomas*. Cancer Treat Res Commun, 2021. **27**: p. 100353.
55. Forest, F., et al., *Heterogeneity of PD-L1 expression in lung adenocarcinoma metastasis is related to histopathological subtypes*. Lung Cancer, 2021. **155**: p. 1-9.
56. 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): p. 1449-1459.
57. Saito, Y., et al., *Inter-tumor heterogeneity of PD-L1 expression in non-small cell lung cancer.* J Thorac Dis, 2019. **11**(12): p. 4982-4991.
58. Planchard, D., et al., *Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2019. **30**(5): p. 863-870.
59. Best, M.G., et al., *RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics.* Cancer Cell, 2015. **28**(5): p. 666-76.
60. Best, M.G., et al., *Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets*. Cancer Cell, 2017. **32**(2): p. 238-252 e9.
61. In 't Veld, S., et al., *Detection and localization of early- and late-stage cancers using platelet RNA*. Cancer Cell, 2022.
62. Blank, C.U., et al., *CANCER IMMUNOLOGY. The "cancer immunogram".* Science, 2016. **352**(6286): p. 658-60.
63. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2016. **165**(1): p. 35-44.
64. McMillan, D.C., *The systemic inflammation-based Glasgow Prognostic Score: a decade of experience in patients with cancer.* Cancer Treat Rev, 2013. **39**(5): p. 534-40.
65. de Vries, R., et al., *Integration of electronic nose technology with spirometry: validation of a new approach for exhaled breath analysis*. J Breath Res, 2015. **9**(4): p. 046001.
66. Ferlay J, S.I., Ervik M, et al. *GLOBOCAN 2012 v1.2, cancer incidence and mortality worldwide: IARC cancerbase no. 11. Lyon: International Agency for Research on Cancer; 2015.* . 2015.
67. Ettinger DS, W.D., Akerley W, et al. . *NCCN Clinical Practice guidelines in oncology. Non-small cell lung cancer. Version 7. 2015*. 2015.
68. Naidoo J, D., A. , *Molecular Diagnostic Testing in Non-Small Cell Lung Cancer.* The american journal of hematology/oncology, 2014. **10**: p. 4-10.
69. Schiller, J.H., et al., *Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer.* N Engl J Med, 2002. **346**(2): p. 92-8.
70. Ardizzoni, A., et al., *Cisplatin- versus carboplatin-based chemotherapy in first-line treatment of advanced non-small-cell lung cancer: an individual patient data meta-analysis.* J Natl Cancer Inst, 2007. **99**(11): p. 847-57.
71. 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, 2008. **26**(21): p. 3543-51.
72. Mok, T.S., et al., *Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma.* N Engl J Med, 2009. **361**(10): p. 947-57.
73. Rosell, R., et al., *Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial.* Lancet Oncol, 2012. **13**(3): p. 239-46.
74. Sequist, L.V., et al., *Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations.* J Clin Oncol, 2013. **31**(27): p. 3327-34.
75. Wu, Y.L., et al., *Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial.* Lancet Oncol, 2014. **15**(2): p. 213-22.
76. Shaw, A.T., et al., *Crizotinib versus chemotherapy in advanced ALK-positive lung cancer.* N Engl J Med, 2013. **368**(25): p. 2385-94.
77. Solomon, B.J., et al., *First-line crizotinib versus chemotherapy in ALK-positive lung cancer.* N Engl J Med, 2014. **371**(23): p. 2167-77.

78. Shaw, A.T., et al., *Ceritinib in ALK-rearranged non-small-cell lung cancer.* N Engl J Med, 2014. **370**(13): p. 1189-97.
79. Janne, P.A., et al., *AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer.* N Engl J Med, 2015. **372**(18): p. 1689-99.
80. Sequist, L.V., L. Rolfe, and A.R. Allen, *Rociletinib in EGFR-Mutated Non-Small-Cell Lung Cancer.* N Engl J Med, 2015. **373**(6): p. 578-9.
81. Seto, T., et al., *CH5424802 (RO5424802) for patients with ALK-rearranged advanced non-small-cell lung cancer (AF-001JP study): a single-arm, open-label, phase 1-2 study.* Lancet Oncol, 2013. **14**(7): p. 590-8.
82. Gadgeel, S.M., et al., *Safety and activity of alectinib against systemic disease and brain metastases in patients with crizotinib-resistant ALK-rearranged non-small-cell lung cancer (AF-002JG): results from the dose-finding portion of a phase 1/2 study.* Lancet Oncol, 2014. **15**(10): p. 1119-28.
83. Shaw, A.T., et al., *Crizotinib in ROS1-rearranged non-small-cell lung cancer.* N Engl J Med, 2014. **371**(21): p. 1963-71.
84. Camidge DR, O.S., Shapiro G, , *Efficacy and safety of crizotinib in patients with advanced c-MET-amplified non-small cell lung cancer (NSCLC).* . J Clin Oncol. , 2014. **32**(Abstract 8001).
85. Planchard D, K.T., Mazierez J, et al. , *LBA38_PR - dabrafenib in patients with BRAF V600E-mutant advanced non-small cell lung cancer (NSCLC): a multicenter, open-label, phase II trial (BRF113928).* Ann Oncol. , 2014. **25**.
86. Planchard D, G.H., Kim TM, et al, *Interim results of a phase II study of the BRAF inhibitor (BRAFi) dabrafenib (D) in combination with the MEK inhibitor trametinib (T) in patients (pts) with BRAF V600E mutated (mut) metastatic non-small cell lung cancer (NSCLC).* . J Clin Oncol. , 2015. **33**.
87. Mazieres, J., et al., *Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives.* J Clin Oncol, 2013. **31**(16): p. 1997-2003.
88. Champiat, S., et al., *Exomics and immunogenics: Bridging mutational load and immune checkpoints efficacy.* Oncoimmunology, 2014. **3**(1): p. e27817.
89. Maemondo, M., et al., *Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR.* N Engl J Med, 2010. **362**(25): p. 2380-8.
90. Yang, J.C., et al., *Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6.* Lancet Oncol, 2015. **16**(7): p. 830-8.
91. Drilon, A., et al., *Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas.* Cancer Discov, 2013. **3**(6): p. 630-5.
92. Pardoll, D.M., *The blockade of immune checkpoints in cancer immunotherapy.* Nat Rev Cancer, 2012. **12**(4): p. 252-64.
93. Ribas, A., *Tumor immunotherapy directed at PD-1.* N Engl J Med, 2012. **366**(26): p. 2517-9.
94. Nishimura, H., et al., *Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor.* Immunity, 1999. **11**(2): p. 141-51.
95. Nishimura, H., et al., *Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice.* Science, 2001. **291**(5502): p. 319-22.
96. Hirano, F., et al., *Blockade of B7-H1 and PD-1 by monoclonal antibodies potentiates cancer therapeutic immunity.* Cancer Res, 2005. **65**(3): p. 1089-96.
97. Wong, R.M., et al., *Programmed death-1 blockade enhances expansion and functional capacity of human melanoma antigen-specific CTLs.* Int Immunol, 2007. **19**(10): p. 1223-34.
98. Curiel, T.J., et al., *Blockade of B7-H1 improves myeloid dendritic cell-mediated antitumor immunity.* Nat Med, 2003. **9**(5): p. 562-7.
99. Nishio M, H.T., Nakagawa K, , *Phase II studies of nivolumab (anti-PD-1, BMS-936558, ONO-4538) in patients with advanced squamous (sq) or nonsquamous (non-sq) non-small cell lung cancer (NSCLC).* J Clin Oncol. , 2015. **33**.
100. Schumacher, T.N. and R.D. Schreiber, *Neoantigens in cancer immunotherapy.* Science, 2015. **348**(6230): p. 69-74.
101. Hussein M, M.M., Chandler J, et al, *Safety and efficacy of nivolumab in an ongoing trial of a PD-L1+/- patient population with metastatic non-small cell lung cancer.* , in *16th World Conference on Lung Cancer;*. 2015: Denver, CO.
102. Carbone DP, S.M., Chen AC, Bhagavatheeswaran P, Reck M, Paz-Ares L, *A phase III, randomized, open-label trial of nivolumab (anti-PD-1; BMS-936558, ONO-4538) versus investigator's choice chemotherapy (ICC) as first-line therapy for stage IV or recurrent PD-L1+ non-small cell lung cancer (NSCLC).* . J Clin Oncol, 2014. **32**.

103. Rizvi N, G.S., Goldman J, et al, *Safety and efficacy of first-line nivolumab (anti-programmed death-1 [PD-1]) and ipilimumab in non-small cell lung cancer (NSCLC).* in *16th World Conference on Lung Cancer*. 2015: Denver, CO.
104. 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): p. 1837-46.
105. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.
106. Wolchok, J.D., et al., *Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria.* Clin Cancer Res, 2009. **15**(23): p. 7412-20.
107. Carbognin, L., et al., *Differential Activity of Nivolumab, Pembrolizumab and MPDL3280A according to the Tumor Expression of Programmed Death-Ligand-1 (PD-L1): Sensitivity Analysis of Trials in Melanoma, Lung and Genitourinary Cancers.* PLoS One, 2015. **10**(6): p. e0130142.
108. McLaughlin, J., et al., *Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer.* JAMA Oncol, 2016. **2**(1): p. 46-54.
109. Meng, X., et al., *Predictive biomarkers in PD-1/PD-L1 checkpoint blockade immunotherapy.* Cancer Treat Rev, 2015. **41**(10): p. 868-76.
110. Reynders, K., et al., *The abscopal effect of local radiotherapy: using immunotherapy to make a rare event clinically relevant.* Cancer Treat Rev, 2015. **41**(6): p. 503-10.
111. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer-associated genes.* Nature, 2013. **499**(7457): p. 214-218.
112. Rizvi, N.A., et al., *Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer.* 2015. **348**(6230): p. 124-128.
113. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
114. Lee, W., et al., *The mutation spectrum revealed by paired genome sequences from a lung cancer patient.* Nature, 2010. **465**(7297): p. 473-7.
115. Govindan, R., et al., *Genomic landscape of non-small cell lung cancer in smokers and never-smokers.* Cell, 2012. **150**(6): p. 1121-34.
116. Ettinger, D.S., et al., *NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 4.2016.* J Natl Compr Canc Netw, 2016. **14**(3): p. 255-64.
117. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* N Engl J Med, 2015. **372**(26): p. 2509-20.
118. Kelderman, S., T.N. Schumacher, and P. Kvistborg, *Mismatch Repair-Deficient Cancers Are Targets for Anti-PD-1 Therapy.* Cancer Cell, 2015. **28**(1): p. 11-3.
119. Weber, J.S., K.C. Kahler, and A. Hauschild, *Management of immune-related adverse events and kinetics of response with ipilimumab.* J Clin Oncol, 2012. **30**(21): p. 2691-7.
120. Long GV, A.V., Ascierto PA, et al. , *Effect of nivolumab (NIVO) on quality of life (QoL) in patients (pts) with treatment-naïve advanced melanoma (MEL): results of a phase III study (CheckMate 066).* J Clin Oncol. . **2015**.
121. Schadendorf D, L.G., Larkin J, et al, *Patient reported outcomes (PROs) from a phase 3 study of nivolumab (NIVO) alone or combined with ipilimumab (IPI) versus IPI in patients with advanced melanoma (MEL): CheckMate 067.*, in *Society for Melanoma Research Congress*. 2015: San Francisco, CA.
122. Abernethy AP, P.M., Chesney JA, et al, *Effect of nivolumab (NIVO) in combination with ipilimumab (IPI) versus IPI alone on quality of life (QoL) in patients (pts) with treatment-naïve advanced melanoma (MEL): results of a phase II study (CheckMate 069).* . J Clin Oncol, 2015. **33**.
123. Reck M, C.C., Taylor F, et al. , *Evaluation of overall health status in patients with advanced squamous non-small cell lung cancer treated with nivolumab or docetaxel in CheckMate 017.* , in *European Cancer Congress*. 2015: Vienna.
124. Grada, A. and K. Weinbrecht, *Next-generation sequencing: methodology and application.* J Invest Dermatol, 2013. **133**(8): p. e11.
125. Strome, S.E., et al., *B7-H1 blockade augments adoptive T-cell immunotherapy for squamous cell carcinoma.* Cancer Res, 2003. **63**(19): p. 6501-5.
126. Vachhani, P. and H. Chen, *Spotlight on pembrolizumab in non-small cell lung cancer: the evidence to date.* Onco Targets Ther, 2016. **9**: p. 5855-5866.
127. Dang, T.O., et al., *Pembrolizumab for the treatment of PD-L1 positive advanced or metastatic non-small cell lung cancer.* Expert Rev Anticancer Ther, 2016. **16**(1): p. 13-20.

128. Lim, S.H., et al., *Pembrolizumab for the treatment of non-small cell lung cancer.* Expert Opin Biol Ther, 2016. **16**(3): p. 397-406.

129. Scapin, G., et al., *Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab.* Nat Struct Mol Biol, 2015. **22**(12): p. 953-8.

130. Keir, M.E., et al., *PD-1 and its ligands in tolerance and immunity.* Annu Rev Immunol, 2008. **26**: p. 677-704.

131. Topalian, S.L., C.G. Drake, and D.M. Pardoll, *Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity.* Curr Opin Immunol, 2012. **24**(2): p. 207-12.

132. Kazandjian, D., et al., *FDA Approval Summary: Nivolumab for the Treatment of Metastatic Non-Small Cell Lung Cancer With Progression On or After Platinum-Based Chemotherapy.* Oncologist, 2016. **21**(5): p. 634-42.

133. Chatterjee, M., et al., *Systematic evaluation of pembrolizumab dosing in patients with advanced non-small-cell lung cancer.* Ann Oncol, 2016. **27**(7): p. 1291-8.

134. 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): p. 255-265.

135. Langer, C.J., et al., *Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study.* Lancet Oncol, 2016. **17**(11): p. 1497-1508.

136. Peters, S., et al., *191TiP: MYSTIC: a global, phase 3 study of durvalumab (MEDI4736) plus tremelimumab combination therapy or durvalumab monotherapy versus platinum-based chemotherapy (CT) in the first-line treatment of patients (pts) with advanced stage IV NSCLC.* Journal of Thoracic Oncology, 2016. **Volume 11**(Issue 4): p. S139-S140.

137. Planchard, D., et al., *A Phase III Study of Durvalumab (MEDI4736) With or Without Tremelimumab for Previously Treated Patients With Advanced NSCLC: Rationale and Protocol Design of the ARCTIC Study.* Clin Lung Cancer, 2016. **17**(3): p. 232-236 e1.

138. Mok, T., et al., *192TiP: NEPTUNE: A global, phase 3 study of durvalumab (MEDI4736) plus tremelimumab combination therapy versus standard of care (SoC) platinum-based chemotherapy in the first-line treatment of patients (pts) with advanced or metastatic NSCLC.* Journal of Thoracic Oncology, 2016. **11**(4): p. S140-S141.

139. Jerusalem, et al., *Javelin in solid tumor: safety and clinical activity of avelumab (anti-PD-L1) as first-line treatment in patients with advanced NSCLC.* , in *WCLC*. 2017: Vienna. p. S252.

140. Brahmer, J.R., et al., *Safety and activity of anti-PD-L1 antibody in patients with advanced cancer.* N Engl J Med, 2012. **366**(26): p. 2455-65.

141. Sunshine, J. and J.M. Taube, *PD-1/PD-L1 inhibitors.* Curr Opin Pharmacol, 2015. **23**: p. 32-8.

142. Patnaik, A., et al., *Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors.* Clin Cancer Res, 2015. **21**(19): p. 4286-93.

143. Khoja, L., et al., *Pembrolizumab.* J Immunother Cancer, 2015. **3**: p. 36.

144. Ahamadi, M., et al., *Model-Based Characterization of the Pharmacokinetics of Pembrolizumab: A Humanized Anti-PD-1 Monoclonal Antibody in Advanced Solid Tumors.* CPT Pharmacometrics Syst Pharmacol, 2017. **6**(1): p. 49-57.

145. Garon EB, R.M., Rodriguez-Abreu D, et al, *Use of a 200 mg fixed dose of pembrolizumab for the treatment of advanced non-small cell lung cancer (NSCLC).* , in *WCLC*. 2016: Vienna.

146. Gadgeel, J.S., Corey J. Langer, Leena Gandhi, Hossein Borghaei, Amita Patnaik, Liza Cosca Villaruz, Matthew A. Gubens, Ralph J. Hauke, James Chih-Hsin Yang, Lecia V. Sequist, Robert D. Bachman, Joy Yang Ge, Harry Raftopoulos, Vassiliki Papadimitrakopoulou, *Pembrolizumab (pembro) plus chemotherapy as front-line therapy for advanced NSCLC: KEYNOTE-021 cohorts A-C.*, in *ASCO*. 2016: Chicago.

147. Gubens LVS, S.J., Powell SF, *Phase I/II study of pembrolizumab (pembro) plus ipilimumab (ipi) as second-line therapy for NSCLC: KEYNOTE-021 cohorts D and H.* , in *ASCO Annual Meeting.* 2016: Chicago.

148. Robert, C., et al., *Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial.* Lancet, 2014. **384**(9948): p. 1109-17.

149. Herbst G, D.-W.B., *Keynote-010: durable clinical benefit in patients with previously treated, PD-L1-expressing NSCLC who completed Pembrolizumab.*, in *WCLC*. 2016: Vienna.

150. Herbst, P.B., Jose Luis Perez-Gracia, Enriqueta Felip, Dong-Wan Kim, Ji-Youn Han, Julian R. Molina, Joo-Hang Kim, Catherine Dubos Arvis, Myung-Ju Ahn, Margarita Majem, Mary J. Fidler, Veerle Surmont, Gilberto Castro, Marcelo Garrido, Yue Shentu, Marisa Dolled-Filhart, Ellie Im, Edward B. Garon, *Archival vs new tumor samples for assessing PD-L1 expression in the KEYNOTE-010 study of pembrolizumab (pembro) vs docetaxel (doce) for previously treated advanced NSCLC.*, in *ASCO annual meeting* 2016: Chicago.
151. 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): p. 147-53.
152. Kim, M.Y., et al., *Clinicopathological analysis of PD-L1 and PD-L2 expression in pulmonary squamous cell carcinoma: Comparison with tumor-infiltrating T cells and the status of oncogenic drivers.* Lung Cancer, 2015. **88**(1): p. 24-33.
153. Mansfield, A.S., et al., *Temporal and spatial discordance of programmed cell death-ligand 1 expression and lymphocyte tumor infiltration between paired primary lesions and brain metastases in lung cancer.* Ann Oncol, 2016. **27**(10): p. 1953-8.
154. Costa, R., et al., *Toxicity profile of approved anti-PD-1 monoclonal antibodies in solid tumors: a systematic review and meta-analysis of randomized clinical trials.* Oncotarget, 2017. **8**(5): p. 8910-8920.
155. Brahmer, J., *Health-related quality of life for pembrolizumab vs chemotherapy in advanced NSCLC with PD-L1 TPS ≥50%: data from KEYNOTE-024* in *WCLC*. 2016, Elsevier: Vienna.
156. EMA. *Summary of opinion on pembrolizumab.* 2016 [cited 2017 March the 28th].
157. Sharyn Katz MH, B.S., Aggarwal C, et al. , *Imaging of Anti-PD1 therapy response in advanced non-small cell lung cancer*, in *WCLC*. 2016: Vienna.
158. Gibney, G.T., L.M. Weiner, and M.B. Atkins, *Predictive biomarkers for checkpoint inhibitor-based immunotherapy.* Lancet Oncol, 2016. **17**(12): p. e542-e551.
159. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.
160. McGranahan, N., et al., *Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade.* Science, 2016. **351**(6280): p. 1463-9.
161. George, J.N., *Platelets.* Lancet, 2000. **355**(9214): p. 1531-9.
162. McAllister, S.S. and R.A. Weinberg, *The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis.* Nat Cell Biol, 2014. **16**(8): p. 717-27.
163. Lefrancais, E., et al., *The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors.* Nature, 2017. **544**(7648): p. 105-109.
164. Denis, M.M., et al., *Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets.* Cell, 2005. **122**(3): p. 379-91.
165. Schubert, S., A.S. Weyrich, and J.W. Rowley, *A tour through the transcriptional landscape of platelets.* Blood, 2014. **124**(4): p. 493-502.
166. Cecchetti, L., et al., *Megakaryocytes differentially sort mRNAs for matrix metalloproteinases and their inhibitors into platelets: a mechanism for regulating synthetic events.* Blood, 2011. **118**(7): p. 1903-11.
167. Rondina, M.T., et al., *The septic milieu triggers expression of spliced tissue factor mRNA in human platelets.* J Thromb Haemost, 2011. **9**(4): p. 748-58.
168. Nilsson, R.J., et al., *Blood platelets contain tumor-derived RNA biomarkers.* Blood, 2011. **118**(13): p. 3680-3.
169. Kirschbaum, M., et al., *Horizontal RNA transfer mediates platelet-induced hepatocyte proliferation.* Blood, 2015. **126**(6): p. 798-806.
170. Michael, J.V., et al., *Platelet microparticles infiltrating solid tumors transfer miRNAs that suppress tumor growth.* Blood, 2017. **130**(5): p. 567-580.
171. Labelle, M., S. Begum, and R.O. Hynes, *Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis.* Cancer Cell, 2011. **20**(5): p. 576-90.
172. Boilard, E., et al., *Platelets amplify inflammation in arthritis via collagen-dependent microparticle production.* Science, 2010. **327**(5965): p. 580-3.
173. Zamora, C., et al., *Binding of Platelets to Lymphocytes: A Potential Anti-Inflammatory Therapy in Rheumatoid Arthritis.* J Immunol, 2017. **198**(8): p. 3099-3108.
174. Schumacher, D., et al., *Platelet-derived nucleotides promote tumor-cell transendothelial migration and metastasis via P2Y2 receptor.* Cancer Cell, 2013. **24**(1): p. 130-7.

175. Labelle, M., S. Begum, and R.O. Hynes, *Platelets guide the formation of early metastatic niches.* Proc Natl Acad Sci U S A, 2014. **111**(30): p. E3053-61.

176. Li, N., *Platelet-lymphocyte cross-talk.* J Leukoc Biol, 2008. **83**(5): p. 1069-78.

177. Rossaint, J., et al., *Directed transport of neutrophil-derived extracellular vesicles enables platelet-mediated innate immune response.* Nat Commun, 2016. **7**: p. 13464.

178. Wang C, S.W., Ye Y, Hu Q, Bomba HN, Gu Z, *In situ activation of platelets with checkpoint inhibitors for post-surgical cancer immunotherapy.* Nat Biomed Eng, 2017. **1**(2): p. 0011.

179. Diem, S., et al., *Neutrophil-to-Lymphocyte ratio (NLR) and Platelet-to-Lymphocyte ratio (PLR) as prognostic markers in patients with non-small cell lung cancer (NSCLC) treated with nivolumab.* Lung Cancer, 2017. **111**: p. 176-181.

180. Nilsson, R.J., et al., *Rearranged EML4-ALK fusion transcripts sequester in circulating blood platelets and enable blood-based crizotinib response monitoring in non-small-cell lung cancer.* Oncotarget, 2016. **7**(1): p. 1066-75.

181. Sol, N., et al., *Tumor-Educated Platelet RNA for the Detection and (Pseudo)progression Monitoring of Glioblastoma.* Cell Rep Med, 2020. **1**(7): p. 100101.

182. Dobbin, K.K. and R.M. Simon, *Sample size planning for developing classifiers using high-dimensional DNA microarray data.* Biostatistics, 2007. **8**(1): p. 101-17.

183. Planchard, D., et al., *Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2018. **29**(Suppl 4): p. iv192-iv237.

184. Schwartz, L.H., et al., *RECIST 1.1-Update and clarification: From the RECIST committee.* Eur J Cancer, 2016. **62**: p. 132-7.

185. Sidorenkov, G., et al., *The OncoLifeS data-biobank for oncology: a comprehensive repository of clinical data, biological samples, and the patient's perspective.* J Transl Med, 2019. **17**(1): p. 374.

186. Best, M.G., et al., *RNA sequencing and swarm intelligence-enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA.* Nat Protoc, 2019. **14**(4): p. 1206-1234.

187. Best, M.G., et al., *RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics.* Cancer Cell, 2015. **28**(5): p. 666-676.

188. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.

189. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

190. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

191. GSEA. *GSEA Human Gene Set: HALLMARK_INFLAMMATORY_RESPONSE* 2022 [cited 2022 November, 2].

192. Hofman, P., et al., *Liquid biopsy in the era of immuno-oncology: is it ready for prime-time use for cancer patients?* Ann Oncol, 2019. **30**(9): p. 1448-1459.

193. Kerr, B.A., et al., *Platelets govern pre-metastatic tumor communication to bone.* Oncogene, 2013. **32**(36): p. 4319-24.

194. Ostrovsky, O., I. Vlodavsky, and A. Nagler, *Mechanism of HPSE Gene SNPs Function: From Normal Processes to Inflammation, Cancerogenesis and Tumor Progression.* Adv Exp Med Biol, 2020. **1221**: p. 231-249.

195. GeneCards. *GeneCards HPSE* [cited 2022 November, 2].

196. Schwenzer, H., et al., *LARP1 isoform expression in human cancer cell lines.* RNA Biol, 2021. **18**(2): p. 237-247.

197. GeneCards. *GeneCards LARP1*. 2022 [cited 2022 November, 2].

198. 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 Thorac Oncol, 2016. **11**(1): p. 39-51.

199. Peters, S., et al., *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): p. 884-896.

200. 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**(17): p. 1675-1684.

201. Sun, R., et al., *A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study.* Lancet Oncol, 2018. **19**(9): p. 1180-1191.

202. 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): p. 843-852 e4.
203. Duruisseaux, M., et al., *Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis.* Lancet Respir Med, 2018. **6**(10): p. 771-781.
204. Cristescu, R., et al., *Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy.* Science, 2018. **362**(6411).
205. 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, randomised phase 3 trial.* Lancet Oncol, 2014. **15**(7): p. 713-21.
206. Tsypin, M., et al., *Extending the information content of the MALDI analysis of biological fluids via multi-million shot analysis.* PLoS One, 2019. **14**(12): p. e0226012.
207. 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**(1): p. 79-86.
208. 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**(1): p. 91.
209. Roder, J., et al., *A dropout-regularized classifier development approach optimized for precision medicine test discovery from omics data.* BMC Bioinformatics, 2019. **20**(1): p. 325.
210. Goodfellow I, Bengio Y, and Courville A, *Deep learning.* Cambridge, 2016: p. 5-11.
211. Breiman, *Out-of-bag estimation.* . 1996.
212. Roder, H., et al., *Robust identification of molecular phenotypes using semi-supervised learning.* BMC Bioinformatics, 2019. **20**(1): p. 273.
213. Weber J, M.A., Roder H, Roder J, Meyer K, Asmellash S, et al, *Pre-treatment patient selection for nivolumab benefit based on serum mass spectra.* J Immunother Cancer 2015(3): p. 103.
214. Grossi F, R.E., Biello F, Barletta G, Maggioni C, Genova C, et al. , *Evaluation of pretreatment serum tests for nivolumab benefit in patients with non-small cell lung cancer.* J Thorac Oncol, 2017. **12**.
215. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
216. Grigorieva J, A.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**: p. 44-53.
217. Roder, J., B. Linstid, and C. Oliveira, *Improving the power of gene set enrichment analyses.* BMC Bioinformatics, 2019. **20**(1): p. 257.
218. Benjamini Y, H., *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* . J Royal Stat Soc, 1995(57): p. 289–300.
219. Goldberg SB, J.L., Kluger HM, Chiang V, Mahajan A, Xia 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**(86).
220. Rutkowski, M.J., et al., *Cancer and the complement cascade.* Mol Cancer Res, 2010. **8**(11): p. 1453-65.
221. Wang, Y., et al., *Autocrine Complement Inhibits IL10-Dependent T-cell-Mediated Antitumor Immunity to Promote Tumor Progression.* Cancer Discov, 2016. **6**(9): p. 1022-35.
222. 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): p. 1385-93.
223. Chang, H.Y., et al., *Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds.* PLoS Biol, 2004. **2**(2): p. E7.
224. Kowanetz M, L.N., Roder J, Oliveira C, Asmellash S, Meyer K, et al, *Evaluation of immune-related markers in the circulating proteome and their association with atezolizumab efficacy in patients with 2L+ NSCLC.* J Immunother Cancer, 2018. **6**: p. 114.
225. Doroshow, D.B. and R.S. Herbst, *Treatment of Advanced Non-Small Cell Lung Cancer in 2018.* JAMA Oncol, 2018. **4**(4): p. 569-570.
226. Hellmann, M.D., et al., *Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden.* N Engl J Med, 2018. **378**(22): p. 2093-2104.
227. van de Kant, K.D., et al., *Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review.* Respir Res, 2012. **13**: p. 117.

228. Wilson, A.D., *Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath.* Metabolites, 2015. **5**(1): p. 140-63.

229. Boots, A.W., et al., *Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises.* Trends Mol Med, 2015. **21**(10): p. 633-644.

230. Phillips, M., T.L. Bauer, and H.I. Pass, *A volatile biomarker in breath predicts lung cancer and pulmonary nodules.* J Breath Res, 2019. **13**(3): p. 036013.

231. Tirzite, M., et al., *Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis.* J Breath Res, 2017. **11**(3): p. 036009.

232. Rocco, G., et al., *Breathprinting and Early Diagnosis of Lung Cancer.* J Thorac Oncol, 2018. **13**(7): p. 883-894.

233. Shlomi, D., et al., *Detection of Lung Cancer and EGFR Mutation by Electronic Nose System.* J Thorac Oncol, 2017. **12**(10): p. 1544-1551.

234. de Vries, R., et al., *Clinical and inflammatory phenotyping by breathomics in chronic airway diseases irrespective of the diagnostic label.* Eur Respir J, 2018. **51**(1).

235. Bossuyt, P.M., et al., *STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.* BMJ, 2015. **351**: p. h5527.

236. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement.* Ann Intern Med, 2015. **162**(1): p. 55-63.

237. De Vries R, M.M., Wolf-Lansdorf M, Baas P, Sterk PJ, Van den Heuvel MM, *Exhaled Breath Analysis for Prediction of Response to Anti-PD1 Therapy in Patients with NSCLC*, in *European Respiratory Society*. 2017.

238. Fujimoto, D., et al., *Predictive Performance of Four Programmed Cell Death Ligand 1 Assay Systems on Nivolumab Response in Previously Treated Patients with Non-Small Cell Lung Cancer.* J Thorac Oncol, 2018. **13**(3): p. 377-386.

239. de Vries, R., et al., *Clinical and inflammatory phenotyping by breathomics in chronic airway diseases irrespective of the diagnostic label.* European Respiratory Journal, 2018. **51**(1).

240. Ahmadzada, T., et al., *An Update on Predictive Biomarkers for Treatment Selection in Non-Small Cell Lung Cancer.* J Clin Med, 2018. **7**(6).

241. Jia, Z., et al., *Detection of Lung Cancer: Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell Lines of Different Histological Origins.* ACS Omega, 2018. **3**(5): p. 5131-5140.

242. Serasanambati, M., et al., *Profiling Single Cancer Cells with Volatolomics Approach.* iScience, 2019. **11**: p. 178-188.

243. Bushdid, C., et al., *Humans can discriminate more than 1 trillion olfactory stimuli.* Science, 2014. **343**(6177): p. 1370-2.

244. Prasad, V. and V. Kaestner, *Nivolumab and pembrolizumab: Monoclonal antibodies against programmed cell death-1 (PD-1) that are interchangeable.* Semin Oncol, 2017. **44**(2): p. 132-135.

245. Reck, M., et al., *Nivolumab plus ipilimumab versus chemotherapy as first-line treatment in advanced non-small-cell lung cancer with high tumour mutational burden: patient-reported outcomes results from the randomised, open-label, phase III CheckMate 227 trial.* Eur J Cancer, 2019. **116**: p. 137-147.

246. Ding, L. and F. Chen, *Predicting Tumor Response to PD-1 Blockade.* N Engl J Med, 2019. **381**(5): p. 477-479.

247. Oken, M.M., et al., *Toxicity and response criteria of the Eastern Cooperative Oncology Group.* Am J Clin Oncol, 1982. **5**(6): p. 649-55.

248. Bassanelli, M., et al., *Heterogeneity of PD-L1 Expression and Relationship with Biology of NSCLC.* Anticancer Res, 2018. **38**(7): p. 3789-3796.

249. Grizzi, G., et al., *Putative predictors of efficacy for immune checkpoint inhibitors in non-small-cell lung cancer: facing the complexity of the immune system.* Expert Rev Mol Diagn, 2017. **17**(12): p. 1055-1069.

250. Goldberg, S.B., et al., *Early Assessment of Lung Cancer Immunotherapy Response via Circulating Tumor DNA.* Clin Cancer Res, 2018. **24**(8): p. 1872-1880.

251. Dromain, C., et al., *Imaging of tumour response to immunotherapy.* Eur Radiol Exp, 2020. **4**(1): p. 2.

252. Hogan, S.A., M.P. Levesque, and P.F. Cheng, *Melanoma Immunotherapy: Next-Generation Biomarkers.* Front Oncol, 2018. **8**: p. 178.

253. Poma JS., G.L., Sanchez JV., D' errico G. , *What do we know about cancer immunotherapy? Long-term survival and immune-related adverse events.* Allergol Immunopathol. (Madri), 2019. **47**(3): p. 303-308.

254. Hirsch, F.R., et al., *PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project.* J Thorac Oncol, 2017. **12**(2): p. 208-222.

255. Gridelli, C., et al., *Predictive biomarkers of immunotherapy for non-small cell lung cancer: results from an Experts Panel Meeting of the Italian Association of Thoracic Oncology.* Transl Lung Cancer Res, 2017. **6**(3): p. 373-386.
256. Yaziji, H. and C.R. Taylor, *PD-L1 Assessment for Targeted Therapy Testing in Cancer: Urgent Need For Realistic Economic and Practice Expectations.* Appl Immunohistochem Mol Morphol, 2017. **25**(1): p. 1-3.
257. Wilson, A.D., *Diverse applications of electronic-nose technologies in agriculture and forestry.* Sensors (Basel), 2013. **13**(2): p. 2295-348.
258. Farraia, M.V., et al., *The electronic nose technology in clinical diagnosis: A systematic review.* Porto Biomed J, 2019. **4**(4): p. e42.
259. Baldini, C., et al., *Electronic Nose as a Novel Method for Diagnosing Cancer: A Systematic Review.* Biosensors (Basel), 2020. **10**(8).
260. de Vries, R., et al., *Prediction of response to anti-PD-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath.* Ann Oncol, 2019.
261. de Vries, R., et al., *Prediction of response to anti-PD-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath.* Ann Oncol, 2019. **30**(10): p. 1660-1666.
262. Paz-Ares, L., et al., *First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial.* Lancet Oncol, 2021. **22**(2): p. 198-211.
263. Bianchi, F., et al., *Solid-phase microextraction coupled to gas chromatography-mass spectrometry followed by multivariate data analysis for the identification of volatile organic compounds as possible biomarkers in lung cancer tissues.* J Pharm Biomed Anal, 2017. **146**: p. 329-333.
264. Ma, W., et al., *Determination of breath gas composition of lung cancer patients using gas chromatography/mass spectrometry with monolithic material sorptive extraction.* Biomed Chromatogr, 2015. **29**(6): p. 961-5.
265. Westhoff, M., et al., *Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study.* Thorax, 2009. **64**(9): p. 744-8.
266. Kort, S., et al., *Multi-centre prospective study on diagnosing subtypes of lung cancer by exhaled-breath analysis.* Lung Cancer, 2018. **125**: p. 223-229.
267. Calandri, M., et al., *The role of radiology in the evaluation of the immunotherapy efficacy.* J Thorac Dis, 2018. **10**(Suppl 13): p. S1438-S1446.
268. Padodara, R. and J. N, *Olfactory Sense in different animals*. The Indian Journal of Vetrenary science, 2014. **2**(1).
269. Jenkins, E.K., M.T. DeChant, and E.B. Perry, *When the Nose Doesn't Know: Canine Olfactory Function Associated With Health, Management, and Potential Links to Microbiota.* Front Vet Sci, 2018. **5**: p. 56.
270. Walker D, W.J., Cavnar P, Taylor J, Pickel D, Hall S, et al. , *Naturalistic quantification of canine olfactory sensitivity.* Appl Anim Behav Sci 2006. **97**: p. 241-254.
271. Bannier, M., et al., *Feasibility and diagnostic accuracy of an electronic nose in children with asthma and cystic fibrosis.* J Breath Res, 2019. **13**(3): p. 036009.
272. Peng, G., et al., *Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors.* Br J Cancer, 2010. **103**(4): p. 542-51.
273. Fernandes, M.P., S. Venkatesh, and B.G. Sudarshan, *Early Detection of Lung Cancer Using Nano-Nose - A Review.* Open Biomed Eng J, 2015. **9**: p. 228-33.
274. Ahmed, W.M., et al., *Exhaled Volatile Organic Compounds of Infection: A Systematic Review.* ACS Infect Dis, 2017. **3**(10): p. 695-710.
275. Guirao, A., et al., *Trained dogs can identify malignant solitary pulmonary nodules in exhaled gas.* Lung Cancer, 2019. **135**: p. 230-233.
276. National Lung Screening Trial Research, T., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* N Engl J Med, 2011. **365**(5): p. 395-409.
277. *ZonMW Nelson study* 2018 [cited 2022 21 november].
278. De Koning HJ, V.D.A.C., ten Haaf K, Oudkerk M. , *Effects of volume CTt lung cancer screening: mortality results of the NELSON randomized controlled population based trial.*, in *IASLC 19th World Conference on Lung Cancer;* . 2018: Toronto.
279. Guirao Montes, L.M.L.-R., Ingrid Ramón Rodríguez, Gemma Sunyer Dequigiovanni, Núria Viñolas Segarra, Ramón María Marrades Sicart, Jorge Hernández Ferrández, Juan José Fibla Alfara, Álvaro Agustí García-Navarro, *Lung cancer diagnosis by trained dogs.* European Journal of Cardio-Thoracic Surgery, 2017. **52**(6): p. 1206-1210.

280. Gion, M., et al., *Circulating tumor markers: a guide to their appropriate clinical use Comparative summary of recommendations from clinical practice guidelines (PART 3).* Int J Biol Markers, 2017. **32**(2): p. e147-e181.

281. Isaksson, S., et al., *CA 19-9 and CA 125 as potential predictors of disease recurrence in resectable lung adenocarcinoma.* PLoS One, 2017. **12**(10): p. e0186284.

282. Holdenrieder, S., et al., *Carcinoembryonic antigen and cytokeratin-19 fragments for assessment of therapy response in non-small cell lung cancer: a systematic review and meta-analysis.* Br J Cancer, 2017. **116**(8): p. 1037-1045.

283. Shirasu, H., et al., *CYFRA 21-1 predicts the efficacy of nivolumab in patients with advanced lung adenocarcinoma.* Tumour Biol, 2018. **40**(2): p. 1010428318760420.

284. Holdenrieder, S., *Biomarkers along the continuum of care in lung cancer.* Scand J Clin Lab Invest Suppl, 2016. **245**: p. S40-5.

285. Harmsma, M., B. Schutte, and F.C. Ramaekers, *Serum markers in small cell lung cancer: opportunities for improvement.* Biochim Biophys Acta, 2013. **1836**(2): p. 255-72.

286. Li, X., et al., *Biomarkers in the lung cancer diagnosis: a clinical perspective.* Neoplasma, 2012. **59**(5): p. 500-7.

287. Holdenrieder, S., et al., *Clinically Meaningful Use of Blood Tumor Markers in Oncology.* Biomed Res Int, 2016. **2016**: p. 9795269.

288. Dal Bello, M.G., et al., *The role of CEA, CYFRA21-1 and NSE in monitoring tumor response to Nivolumab in advanced non-small cell lung cancer (NSCLC) patients.* J Transl Med, 2019. **17**(1): p. 74.

289. Moritz, R., et al., *Diagnostic validation and interpretation of longitudinal circulating biomarkers using a biomarker response characteristic plot.* Clin Chim Acta, 2018. **487**: p. 6-14.

290. Corcoran, R.B. and B.A. Chabner, *Application of Cell-free DNA Analysis to Cancer Treatment.* N Engl J Med, 2018. **379**(18): p. 1754-1765.

291. Wan, J.C.M., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA.* Nat Rev Cancer, 2017. **17**(4): p. 223-238.

292. Brozos-Vazquez, E.M., et al., *Immunotherapy in nonsmall-cell lung cancer: current status and future prospects for liquid biopsy.* Cancer Immunol Immunother, 2020.

293. Noonan, S.A., et al., *Baseline and On-Treatment Characteristics of Serum Tumor Markers in Stage IV Oncogene-Addicted Adenocarcinoma of the Lung.* J Thorac Oncol, 2018. **13**(1): p. 134-138.

294. Nederland, Z.i., *Het farmacotherapeutisch kompas, referentiewaarden*. https://www.farmacotherapeutischkompas.nl/farmacologie/referentiewaarden. Page seen at 18th of october 2019.

295. Husain, H., et al., *Monitoring Daily Dynamics of Early Tumor Response to Targeted Therapy by Detecting Circulating Tumor DNA in Urine.* Clin Cancer Res, 2017. **23**(16): p. 4716-4723.

296. Phallen, J., et al., *Early Noninvasive Detection of Response to Targeted Therapy in Non-Small Cell Lung Cancer.* Cancer Res, 2019. **79**(6): p. 1204-1213.

297. Trape, J., et al., *Increased plasma concentrations of tumour markers in the absence of neoplasia.* Clin Chem Lab Med, 2011. **49**(10): p. 1605-20.

298. Sevinc, A., et al., *How to interpret serum CA 125 levels in patients with serosal involvement? A clinical dilemma.* Oncology, 2003. **65**(1): p. 1-6.

299. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.

300. Haanen, J., et al., *Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2017. **28**(suppl_4): p. iv119-iv142.

301. Martins, F., et al., *Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance.* Nat Rev Clin Oncol, 2019. **16**(9): p. 563-580.

302. Larsen KG, P.P., Yi W., *UPPAAL in a nutshell.* Int J Softw Tools Technol Transfer, 1997. **1**((1-2)): p. 134-152.

303. Alur R, D.D., *The Theory of Timed Automata*. 1992, Berlin: Springer.

304. David A, L.K., Legay A, Mikučionis M, Poulsen DB, *Uppaal SMC tutorial.* Int J Softw Tools Technol Transfer, 2015. **17**(4): p. 397-415.

305. van Delft F, L.R., *Simulating the diagnostic protocol aimed at detection of immunotherapy related adverse events: UPPAAL model description.* 2020.

306. Institute, N.C. *Common Terminology Criteria for Adverse Events (CTCAE).* . 2020 March 27, 2020 [cited July, 2 2020].

307. Team, R.C. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing 2019.
308. RC., T. *Statistical Computing.* R Foundation for Statistical Computing 2019.
309. H., W., *ggplot2: Elegant Graphics for Data Analysis*. 2016, New York: Springer-Verlag.
310. Massart, *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality.* Annals of Probability, 1990. **18**(3): p. 1269-1283.
311. Gowen, M.F., et al., *Baseline antibody profiles predict toxicity in melanoma patients treated with immune checkpoint inhibitors.* J Transl Med, 2018. **16**(1): p. 82.
312. Maekura, T., et al., *Predictive Factors of Nivolumab-induced Hypothyroidism in Patients with Non-small Cell Lung Cancer.* In Vivo, 2017. **31**(5): p. 1035-1039.
313. Peng, L., et al., *Peripheral blood markers predictive of outcome and immune-related adverse events in advanced non-small cell lung cancer treated with PD-1 inhibitors.* Cancer Immunol Immunother, 2020. **69**(9): p. 1813-1822.
314. von Itzstein MS, K.S., Gerber DE. , *Investigational biomarkers for checkpoint inhibitor immune-related adverse event prediction and diagnosis.* Clin Chem, 2020. **66**(6): p. 779-793.
315. Andre, T., et al., *Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer.* N Engl J Med, 2020. **383**(23): p. 2207-2218.
316. Subbiah, V., et al., *The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) >/=10: a decision centered on empowering patients and their physicians.* Ann Oncol, 2020. **31**(9): p. 1115-1118.
317. Muller, M., et al., *Pembrolizumab for the treatment of non-small cell lung cancer.* Expert Rev Anticancer Ther, 2017. **17**(5): p. 399-409.
318. Marabelle, A., et al., *Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study.* Lancet Oncol, 2020. **21**(10): p. 1353-1365.
319. Pfeifer, G.P., et al., *Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers.* Oncogene, 2002. **21**(48): p. 7435-51.
320. Li, W.Q., et al., *Cost-effectiveness of programmed cell death ligand 1 testing and tumor mutational burden testing of immune checkpoint inhibitors for advanced non-small cell lung cancer.* Chin Med J (Engl), 2020. **133**(21): p. 2630-2632.
321. Meri-Abad, M., et al., *Clinical and technical insights of tumour mutational burden in non-small cell lung cancer.* Crit Rev Oncol Hematol, 2023. **182**: p. 103891.
322. Wiley. *Levels of Evidence*. 2023 [cited 2023 14 april].
323. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
324. Schouten, R.D., et al., *Clinical Utility of Plasma-Based Comprehensive Molecular Profiling in Advanced Non-Small-Cell Lung Cancer.* JCO Precis Oncol, 2021. **5**.
325. Callaham, M.L., et al., *Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting.* JAMA, 1998. **280**(3): p. 254-7.
326. de vries, r., *Binnen een minuut uitslag.* . 2022: https://www.folia.nl/wetenschap/153485/binnen-een-minuut-uitslag-uvaer-promoveert-op-nepneus-die-longziekten-herkent.
327. UMC, A. *Miljoenen van KWF voor kankeronderzoek bij Amsterdam UMC*. 2022 21th of december, 2022 [cited 2022 27th of december].
328. Ayoub, N.M., *Editorial: Novel Combination Therapies for the Treatment of Solid Cancers.* Front Oncol, 2021. **11**: p. 708943.
329. Socinski, M.A., et al., *Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC.* N Engl J Med, 2018. **378**(24): p. 2288-2301.
330. Dong, Q., et al., *Host-Microbiome Interaction in Lung Cancer.* Front Immunol, 2021. **12**: p. 679829.
331. Routy, B., et al., *Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors.* Science, 2018. **359**(6371): p. 91-97.
332. Adachi, K. and K. Tamada, *Microbial biomarkers for immune checkpoint blockade therapy against cancer.* J Gastroenterol, 2018. **53**(9): p. 999-1005.
333. Ocariz-Diez, M., et al., *Microbiota and Lung Cancer. Opportunities and Challenges for Improving Immunotherapy Efficacy.* Front Oncol, 2020. **10**: p. 568939.