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

Muller, M.

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.