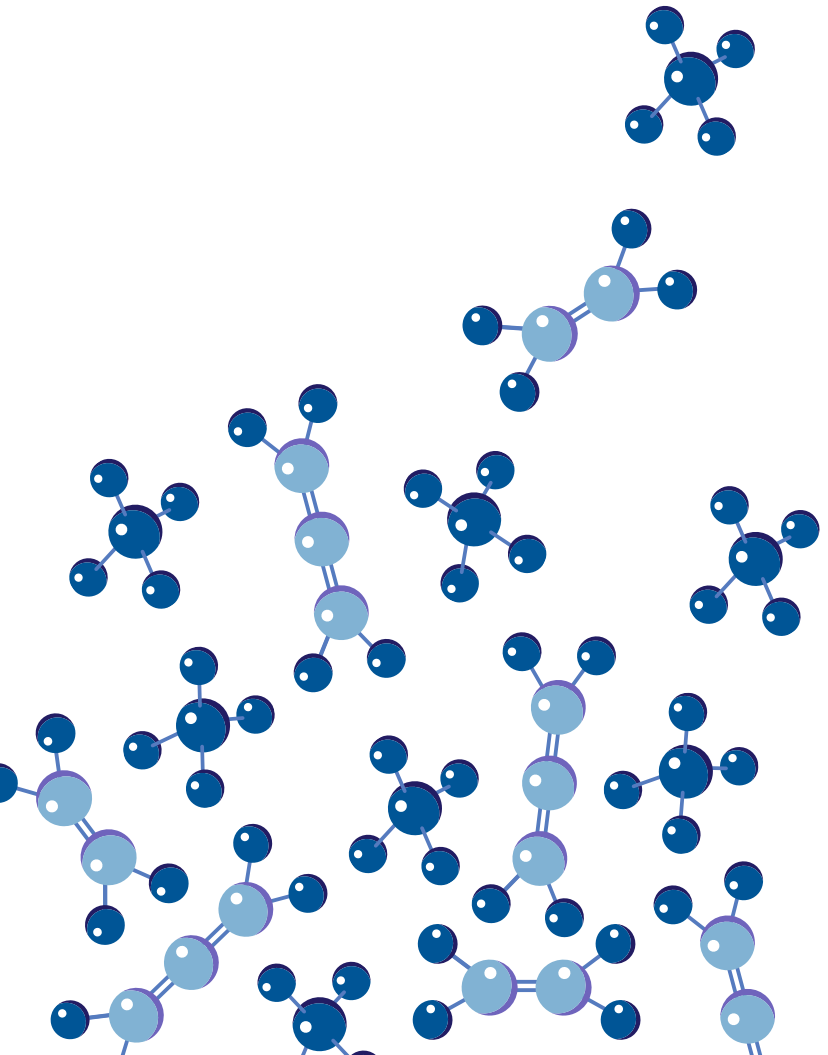
Biomarkers for the response to immunotherapy in patients with non-small cell lung cancer

Muller, M.



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).

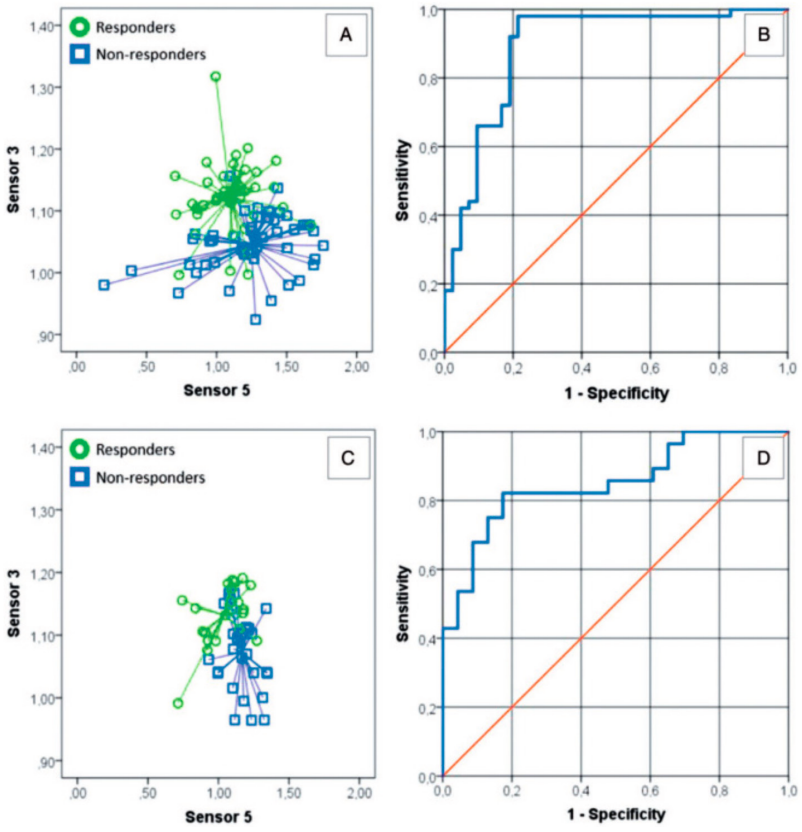


Figure 1 - Results

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].

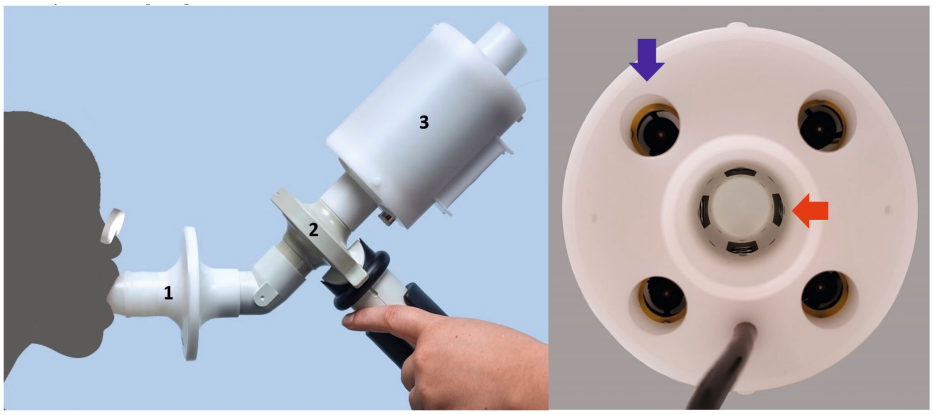


Figure S1 - SpiroNose measurement setup
Left: (1) Mouthpiece, nose clamp and bacteria filter, (2) Spirometer, (3) SpiroNose. Right: Front view of the SpiroNose and the positioning of the sensor arrays. Red arrow: 4 sensor arrays monitoring exhaled breath. Blue arrow: 4 reference sensor arrays monitoring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

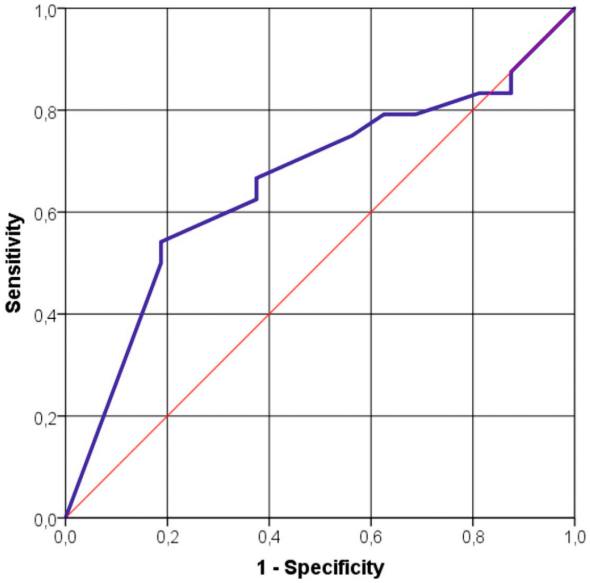


Figure S3 - Validation set: ROC curve with line of identity of the PD-L1 immunohistochemistry predictive for the selection of non-responders to anti-PD-1 therapy (AUC: 0.66). PD: progressive disease, PR: partial response, SD: stable disease, RECIST: Response Evaluation Criteria in Solid Tumors, PD-L1: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

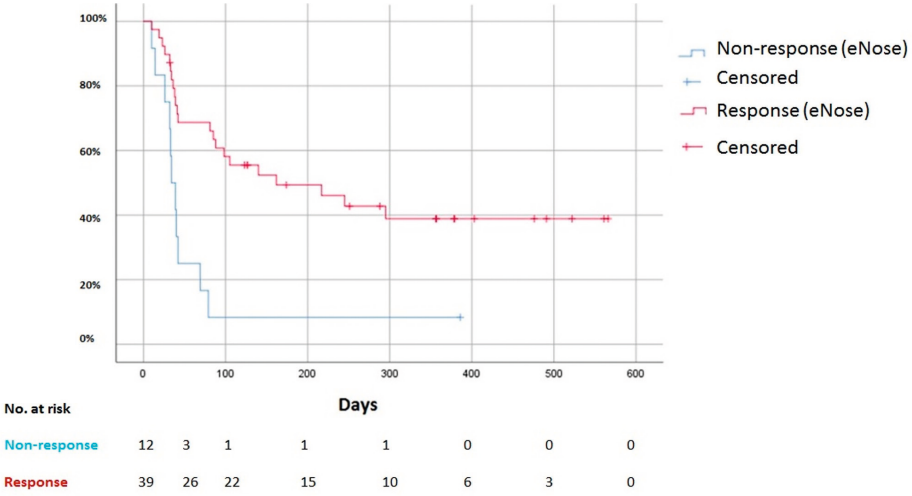


Figure S4A – Progression Free Survival: In this figure the progression free survival (PFS) analysis of the two groups based on the eNose test (with a cut-off of 0.72) of the validation set is shown with a Kaplan Meier curve. In the non-responder group according to the eNose test, one patient is censored (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).

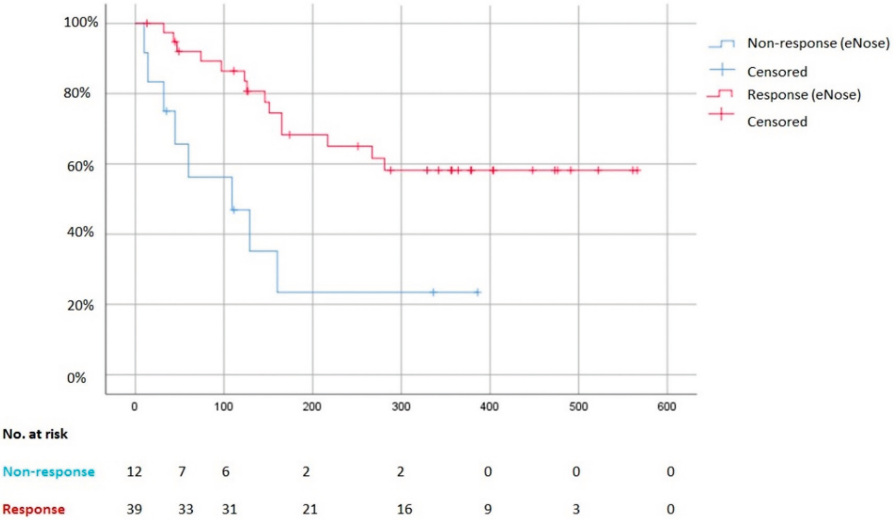


Figure S4B – Overall Survival: In this figure the overall survival (OS) analysis of the two groups based on the eNose test (with a cut-off of 0.72) of the validation set is shown with a Kaplan Meier curve. The median overall survival was 109 days (IQR: 32 - 160 days) for the non-responder group and not 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).