Aus der Medizinischen Klinik und Poliklinik III Großhadern
Klinikum der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. Dr. Michael von Bergwelt

VOC Pattern Recognition of Lung Cancer: a Comparative Evaluation of Canine and eNose- Based Strategies Using Different Sampling Materials

Dissertation
zum Erwerb des Doktorgrades der Humanmedizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Wiebke Ingalisa Biehl
aus
Chivhu, Zimbabwe
2021

Mit Genehmigung der Medizinischen Fakultät

der Universität München

Berichterstatter	: Prof. Dr. rer. nat. Helga Schmetzer
Mitberichterstatter	: PD Dr. med. Katrin Kahnert PD Dr. Claudia Staab-Wiejnitz, PhD apl. Prof. Dr.rer. boil. hum. Katja Radon
Dekan	: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 25.02.2021

Inhaltsverzeichnis

1. Zusammenfassung

Hintergrund: Lungenkrebs ist der häufigste Grund für krebsbedingte Todesfälle. Selbst bei optimaler Behandlung versterben 80-90 % der an Lungentumoren erkrankten Patienten innerhalb von 5 Jahren. Die Prognose kann allerdings signifikant verbessert werden, wenn die Erkrankung in frühen Stadien erkannt und behandelt wird. In den letzten Jahren wurde berichtet, dass Hunde in der Lage sind (Lungen-) Tumore zu erriechen. Ferner wurde berichtet, dass eine elektronische Nase (eNose), welche chemische Sensoren für die Detektion von VOC beinhaltet und ebenfalls über eine Geruchsmustererkennung arbeitet, in der Lage ist, volatile organische Komponenten (VOC) zu detektieren. Beides könnte im Rahmen eines Screening-Tests dazu beitragen Tumore in frühen Stadien zu erkennen.

Ziele: Im Rahmen dieser Arbeit sollten folgende Fragestellungen bearbeitet werden:
1. Ein Vergleich der Ergebnisse von erfahrenen Arbeitshunden und reinen Familienhunden bezüglich der Erkennung von Exhalat von gesunden Probanden und Patienten mit nicht malignen Lungenerkrankungen (Spezifität) und Exhalat von Tumorpatienten (Sensitivität), um anschließend ein optimiertes Training zu entwickeln
2. Der Einfluss unterschiedlicher untersuchter Exhalat-bindender Trägermaterialien auf die von den Hunden erzielten Ergebnisse
3. Ein Vergleich von in Exhalatbeutel gesammeltem Exhalat mit auf Trägermaterialien gesammeltem Exhalat mit Hilfe einer elektronischen Nase
4. Ein Vergleich zwischen den von Hunden erzielten Ergebnissen und den Ergebnissen der elektronischen Nase Cyranose 320 ™
5. Ableitung einer Strategie zur Überprüfung von VOC mittels Hunden und elektronischer Nase mittels eines geeigneten Trägermaterials

Material und Methoden: Im ersten Teil der Studie, dem methodischen Ansatz, wurden zwei Hundeteams eingeschlossen. Hundeteam 1 arbeitete mit fünf erfahrenen Arbeitshunden, während Hundeteam 2 fünf gewöhnliche Familienhunde ohne vorherige Arbeitserfahrung ausbildete, um eine Aussage zur Qualifikation der verschiedenen Hunde sowie zur besten Ausbildungsmethode treffen zu können. Um das beste Trägermaterial für Exhalat zu finden, verglichen wir mit Aktivkohle gefüllte Glasröhrchen und Mundschutze aus Vlies, die in Plastikbechern gelagert wurden. Es

wurden 70 Tumoratemproben sowie 88 Kontrollatemproben von Patienten ohne maligne Lungenerkrankungen und von gesunden Probanden in der Asklepios Klinik Gauting abgenommen.

Im zweiten Teil der Studie wurden 5 erfahrene Arbeitshunde ausgewählt, welche nach den Erfahrungen des ersten Ansatzes mit überarbeiteten und verbesserten Trainingsmethoden ausgebildet wurden. Zur Abnahme des Exhalats wurden zwei Trägermaterialien auf Vliesbasis gewählt: zum einen Glasröhrchen mit zwei unterschiedlichen (silikonisierten bzw. nicht silikonisierten) Vliesstoffen, die von der Hundegruppe genutzt wurde, zum anderen die zuvor getesteten Vlies-Mundschutze. Beide Trägermaterialien wurden mit der elektronischen Nase getestet. Exhalat von 9 Patienten mit Lungentumoren sowie Exhalat von 35 COPD-Patienten und gesunden Probanden wurde abgenommen.

Ergebnisse: Insgesamt zeigte sich, dass der Einsatz von erfahrenen Arbeitshunden dem von Familienhunden im ersten Studienteil überlegen war. Die Sensitivität der Hunde des ersten Studienteils für die Erkennung von Exhalat von Tumorpatienten lag bei 45-59 %, die Spezifität bei 45-69 %. Auf Aktivkohle basierende Trägermaterialien waren für die Erkennung von volatilen organischen Substanzen durch Hunde ungeeignet.

Die erzielte Spezifität der Hunde im zweiten Teil betrug 83 %, die Sensitivität lag bei 56 %, allerdings mit erheblichen Unterschieden zwischen den einzelnen Hunden. Die elektronische Nase erbrachte für beide vliesbasierten Trägermaterialien eine Spezifität von 97 %, eine Sensitivität von 89 % für Vliesstoffe in Glasröhrchen und 100 % Sensitivität für Vlies-Mundschutze. Messungen von direkt von Patienten in Exhalatbeuteln gesammeltem Exhalat als Referenzmessungen erzielte eine Sensitivität und Spezifität von 100 %.

Fazit: Die Daten zeigen grundsätzlich, dass sowohl erfahrene Arbeitshunde als auch reine Familienhunde das Potenzial haben, Exhalat von Patienten mit Tumorerkrankung von Exhalat von Patienten und Probanden ohne maligne Lungenerkrankungen zu unterscheiden. Erfahrene Arbeitshunde waren leichter und schneller auszubilden und erzielten bessere Ergebnisse. Dennoch war die Sensitivität und Spezifität der Hundeergebnisse sehr stark von der Art des Trainings, der

individuellen Leistungsfähigkeit des einzelnen Hundes und des genutzten Trägermaterials abhängig. Ein Vergleich der Hundeergebnisse mit denen einer elektronischen Nase zeigte bessere Ergebnisse sowohl für die Spezifität als auch für die Sensitivität für die elektronische Nase.

Sowohl Vliesstoffe in Glasröhrchen als auch vliesbasierte Mundschutze können erfolgreich als Trägermaterialien für Exhalat genutzt werden. Die Möglichkeit, Exhalat mit geeigneten Trägermaterialien abzunehmen, eine bestimmte Zeit zu lagern und an einen Ort verschicken zu können, an welchem sie dann mit einer elektronischen Nase ausgewertet werden können, bietet viele Einsatzmöglichkeiten für weitere Studien.

2. Abstract

Parts of the abstract are already published in Biehl et al., Acta oncologica 2019 [1].

Background: Lung cancer is the leading cause of cancer related death. Even with optimal treatment, 80-90% of lung cancer patients die within 5 years. However, the prognosis can be significantly improved if the disease is detected and treated in early stages. In recent years, it has been reported that dogs are able to detect (lung) cancer. It has also been reported that an electronic nose (eNose) with chemical sensors can detect volatile organic compounds (VOC) via VOC pattern recognition. Both, dogs and eNose, could help to identify tumors in their early stages by screening.

Aims: In this study the following analysis were performed:
1. Comparison of the results of experienced working dogs versus family dogs regarding the detection capability of non-cancer breath samples (specificity) and cancer breath samples (sensitivity) in order to understand how to optimize training.
2. Influence of different breath sample carrier materials on the results achieved by the dogs.
3. Comparison of breath samples by eNose, for those collected and directly assessed in respiratory bags, with those collected on carrier materials and assessed at a later time.
4. Comparison of results achieved by dogs with results by eNose.
5. Development of a strategy for a volatile profiling by dogs and eNose using a suitable carrier material.

Material and methods: In the first part of the study, using a methodological approach, two dog teams were employed. Dog Team 1 worked with 5 experienced working dogs, while dog team 2 trained 5 ordinary family dogs with no prior work experience, to discover which dogs were better qualified and the best training method.

To find the best carrier material for breath sampling, we compared charcoal filled glass tubes with fleece based earloop masks stored in plastic cups. Breath samples were collected at the Asklepios Klinik Gauting; 70 cancer breath samples from patients with malignant lung disease, and 88 control breath samples from healthy subjects.

In the second part of the study, 5 experienced working dogs were trained with revised and improved training methods learning from experiences in the first part. Two fleece-based carrier materials were selected for breath sample collection: a) glass tubes containing two different (siliconized and non-siliconized) fleeces and the previously tested fleece earloop masks. Testing was done by the dog group on the fleeces in glass tubes and by eNose on both breath sample carrier materials. 9 breath samples from patients with lung cancer, as well as 35 control breath samples from COPD patients and healthy volunteers were taken.

Results: In the first part of the study it was shown overall that experienced working dogs performed better than family dogs and the dogs achieved a sensitivity of 45-59% and a specificity of 45-69%. Charcoal based breath sample carrier materials did not qualify for detection of VOC by dogs. In the second part of the study, the dogs achieved a specificity of 83% and a sensitivity of 56%, but with considerable differences between individual dogs. The eNose provided a specificity of 97% for both fleece based carrier materials and a sensitivity of 89% for fleece filled glass tubes and 100% for earloop masks. Measurements of breath samples collected directly in respiratory bags as reference measurements achieved a sensitivity and specificity of 100%.

Conclusion: Our data shows that both experienced working dogs as well as family dogs have the potential to distinguish between breath samples from cancer patients and non-cancer samples. Experienced working dogs can be trained more easily, faster, and achieved better results. However, the accuracy of the dogs depended very much on; the type of training; the performance of the individual dog; and the carrier material used. A comparison of the dogs' results to those of eNose showed better results for both specificity and sensitivity by eNose.

Both tested carrier materials, fleeces in glass tubes and fleece based earloop masks, can be successfully used as carrier materials for breath samples. There are many possibilities for further eNose studies such as collecting breath samples with qualified carrier material, storing them for a certain period of time, and sending them to a location where they can then be assessed by eNose.

3. Introduction

Parts of the introduction are already published in Biehl et al., Acta oncologica 2019 [1].

Lung cancer (LC) is the leading cause of cancer-related death [2] and often occurs at the site of damaged lung tissue and inflammation, especially in smokers with or without a history of chronic obstructive pulmonary disease (COPD) [3]. Even if treated optimally, 80-90% of LC patients die within 5 years after initial diagnosis. The prognosis can be significantly improved, if the disease is detected and treated in early stages, resulting in survival rates of up to 70% [4]. Diagnostic procedures such as standard chest X-ray, sputum cytology or computer tomography (CT) in combination with positron emission tomography show high rates of both false negative and false positive results, leading to expensive diagnostic procedures and unwarranted surgery [5]. In the last years many attempts have been made to develop more effective screening methods for LC, such as (PCR-based) sputum analysis, CT image analysis, or fluorescence bronchoscopy [6-10]. For asymptomatic patients with a high risk for LC, particularly active, or former smokers~~,~~ who ceased smoking within 15 years, with at least 30 pack years and of age 55-80 years, the U.S. Preventive Services Task Force recommends low-dose CT (LDCT) once a year. Patients seem to benefit from LDCT screening, but in those with low LC risk, overdiagnosis or the risk originating from radiation might dominate [11].

Several tests have been proposed that are based on biomarkers contained in exhaled air, which can be collected simply and non-invasively. The human breath contains >3000 different substances in terms of volatile organic compounds (VOC) [12]. Even if produced in distant organs, VOC are found in the exhaled air, as they are transported to the lung via the blood stream. Different studies have shown that the profiles of VOC in the exhaled air are associated with diseases, although only rarely specific, VOC profiles could be linked to specific alterations in organ function. Investigators mostly focused on the discrimination of the VOC patterns of healthy subjects and cancer patients, using gas chromatography, mass spectrometry, electronic nose (eNose), or colorimetric sensor arrays. For LC patients the analysis of up to 22 VOC found that these methods could discriminate control from cancer samples with specificities of 67-99% and sensitivities of 54-86%, whereby the

differences in results were thought to be due to the lack of standardized sampling as well as different statistical strategies [13].

Results obtained with an eNose, containing chemical sensors for the detection of VOC profiles ('breathprints') in combination with algorithms for pattern recognition, have shown that in general an eNose is capable of discriminating breath samples from patients with (lung) cancer or healthy subjects and even to obtain individual 'breathprints'. Several studies with different types of eNoses have confirmed the possibility of discrimination between patients with LC from controls subjects without LC or from healthy subjects, with promising results for sensitivity and specificity [14-18]. However, the available eNoses are not sensitive enough to detect or identify individual VOCs in low concentrations [19].

In recent years, it has also been shown that specially trained sniffer dogs can differentiate between samples from cancer patients with samples from healthy subjects and patients with non-malignant disease. While some studies worked with samples from patients with ovarial or breast cancer, colorectal cancer or prostate cancer [20-23], most approaches have been made with breath samples from LC patients [21;23-26]. Also, different approaches have studied sample carrier materials and collection methods. While some studies used tumor tissue, urine or stool samples [20-22;25], most of the other studies worked with breath samples, collected with different carrier materials, such as sampling tubes filled with different fleeces or charcoal or sterile exhalation filters [21;23-26]. Detection sensitivity and specificity differed depending on the experimental setup. Including single cancer samples among a number of control samples in a blind test yielded a 71-99% sensitivity and a 91-99% specificity, usually obtained as a 'collective decision' of all participating dogs [20-24]. When the dogs were confronted with a situation of variable numbers of cancer samples (0-6 cancer samples in each trial), sensitivity and specificity decreased to 56% and 34% respectively [25-26]. While several studies have trained household dogs with only basic obedience training, other studies used dogs with working-experience like scent tracking or search and rescue dogs.

Based on these studies and their results the aims of the study were:

1) To figure out in a first methodological approach, potential differences in sensitivity or specificity achieved by experienced working dogs (that had already been trained for cancer breath detection) and those by ordinary household dogs. In addition, training methods of two participating dog teams would be compared to find the most qualifying training strategy.

2) Hypothesizing that VOC can be provided by carrier materials in the methodological approach, two different carrier materials for breath sample collection and their possible influences on dogs' results would be compared: One dog team used charcoal in glass tubes (Hackner et al) [26], the other group simple fleece earloop masks.

3) In a second, validating part, we would work with experienced sniffing working dogs trained with an optimized training strategy (developed during the methodological approach). Also, only a fleece-based breath sample carrier material (filled in glass tubes) would be used for dogs testing (McCulloch et al., Ehmann et al.) [23-24].

4) To compare different fleece-based carrier materials and, also to have an objective method of VOC analysis for reference, an additional VOC analysis with an eNose (Cyranose 320 TM) was initiated. Results obtained by the dogs and the eNose would be compared.

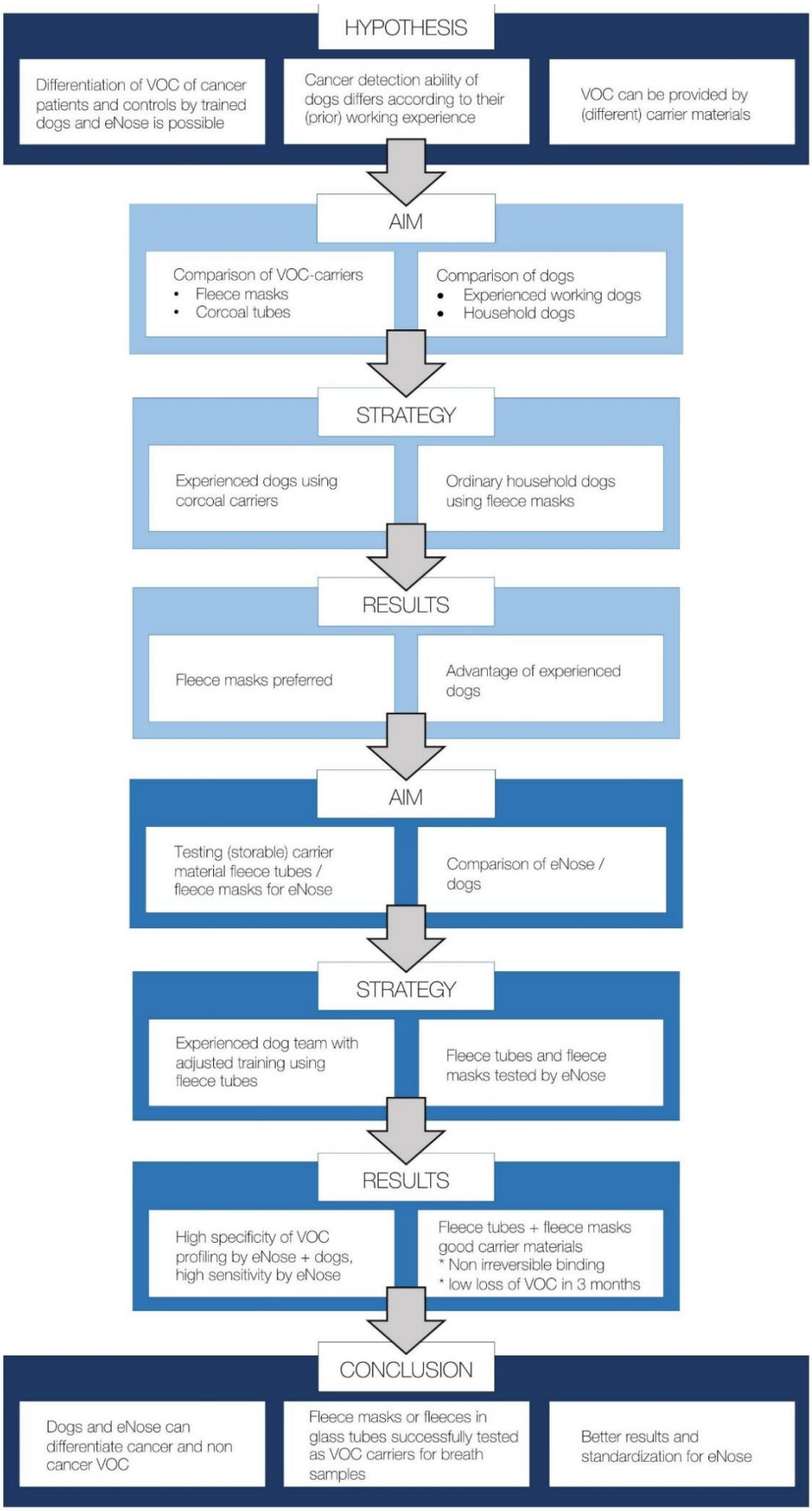

Figure 1. Overview total study. Overview of the total study incl. hypothesis, aims, strategy, results and conclusion.

4. Material and Methods

Parts of the material and methods are already published in Biehl et al., Acta oncologica 2019 [1].

4.1 Methodological part of the study

In a preliminary methodological approach results of experienced working dogs (that had already been trained for cancer breath detection) and ordinary household dogs were compared and different conditioning strategies were chosen for the dogs' training. Also, two different carrier materials were used by the two dog teams.

4.1.1 Patients and healthy subjects

The study protocol to collect breath samples from patients and subjects with and without LC was approved by the local ethical committee and all participants gave their written informed consent. Participants' eating, smoking or other consumption habits as well as concomitant diseases or medications were collected in order to collect information about factors that could influence the VOC composition of exhaled breath. In addition, the medical record was used to collect further information. Cancer patients were included irrespective of their stages of cancer, either before chemotherapy (if chemotherapy was planned), or at least one year after finishing chemotherapy for residual tumor. Breath samples were collected in the Asklepios Klinik Gauting and sampling was performed in rooms of the hospital to assure a comparable 'background smell'.

We included 71 patients with malignant cancer in the lung: small cell LC (SCLC, n=7); non-small cell LC (NSCLC): adenocarcinoma (n= 21), squamous cell carcinoma (n=17); tumors of the lung with other histologies (n=15) or other solid tumors metastasized to the lung (n=11). Patients with non-malignant lung diseases were included as control group (COPD and other respiratory diseases, interstitial lung disease, pleura diseases, pneumonia and other lung non-malignant lung diseases (n=47)). Subjects without lung-diseases served as healthy controls (n=43). 75 male and 86 female patients and healthy subjects were enrolled in the trial (male:female ratio was 1:0.87). The average subject ages were respectively; cancer patients - 64

years (range 43-84); patients with non-malignant lung diseases - 64 years (range 18-86); and healthy subjects 54 years (range 23-85).

Table 1. Patients' characteristics methodological approach

Diagnosis	Subtype	Gender [n] (m/f)[1]	Age [Ø y][1] (range)	Stage of disease[2] [n]				
Methodological Approach				I	II	III	IV	n.a.
Lung cancer (n=71)		37/34	64 (43-84)	9	10	13	13	25
	SCLC[3]	3/4				4	2	1
	NSCLC[4]	11/10		3	4	5	5	4
	Adenocarcinoma							
	Squamous	11/6		3	5	2	4	3
	Other	8/7		2	1	2	1	9
	Lung metastasis	4/7						
Non-malignant lung disease (n=47)	COPD[5]	14/12	64 (18-86)		7	7	8	2
	Other	6/15						
Healthy subjects (n=43)		18/25	54 (23-85)					

[1] m male; f female; y years; [2] based on the TNM-classification for Lung cancer; GOLD criteria for classifying stages of COPD; [3] Small-cell lung cancer (SCLC); [4] Non small-cell lung cancer (NSCLC); [5] Chronic obstructive pulmonary disease (COPD)

4.1.2 Breath sample collection

The breath samples of each participant were collected with two different sampling materials. One was a charcoal-filled glass tube (Draeger, Lübeck, Germany). Participants' exhailed air was collected in a bag and then slowly drawn through the charcoal-filled glass tube, similar to the method used by Hackner et al. [26]. The other sampling material was a fleece earloop mask (Henry Schein Medical GmbH, Hamburg, Germany). Participants had to breathe through the mask for 3-5 min - the fleece was stored in plastic cups with a lid afterwards. In all sample collections, both experimenters and sample donors were required to wear unpowdered Latex gloves (Meditrade). Participants were required to fast and asked not to brush their teeth or to use any cosmetics (like eau de toilette, after shave balm and lipstick) for at least 2 h before sampling. All samples were put in zipper-bags and stored in the dark at 10°C until further use in the following weeks. A small number of breath samples were duplicates of the same patient and presented to the dogs at different times in order to check reproducibility of the dogs' results over time and thus the quality of the samples after storage.

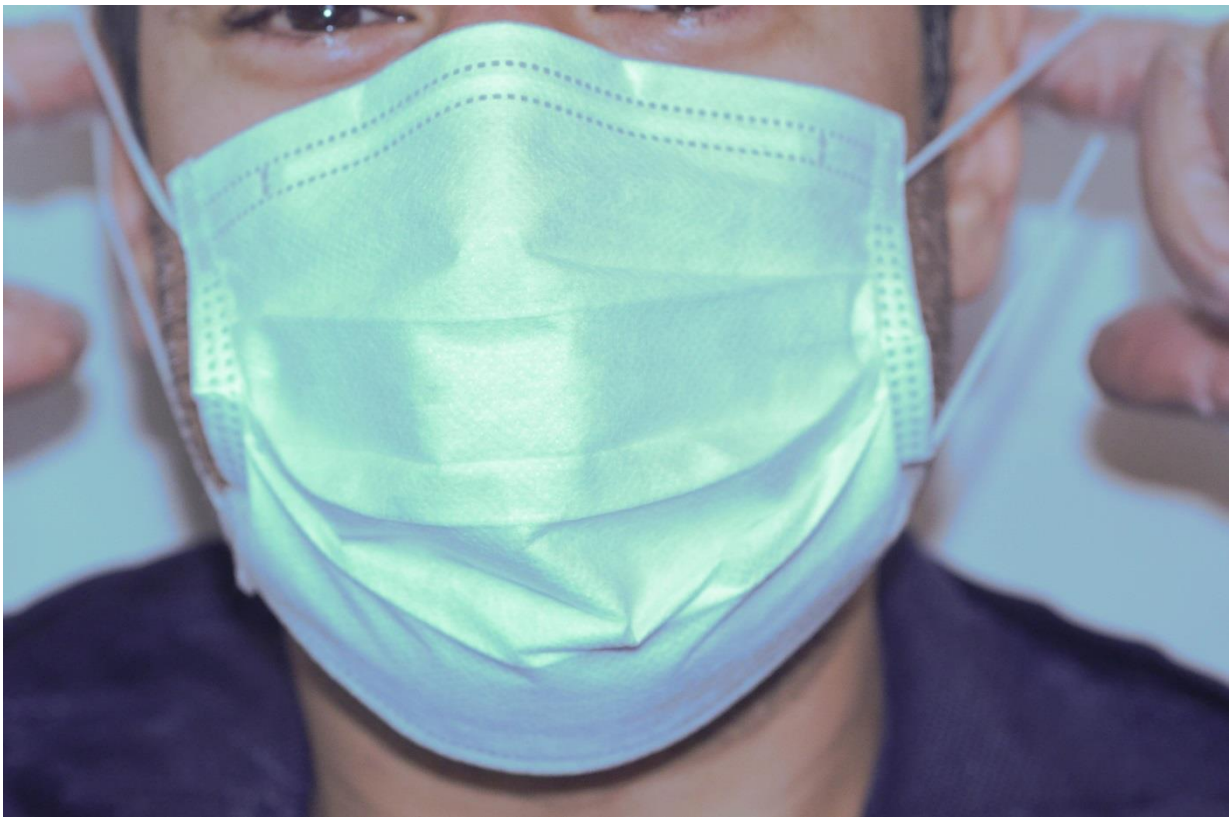

Figure 2. Sample collection with earloop mask. Breath sample collection with earloop masks. Participants had to breathe through earloop masks for 3-5 minutes.

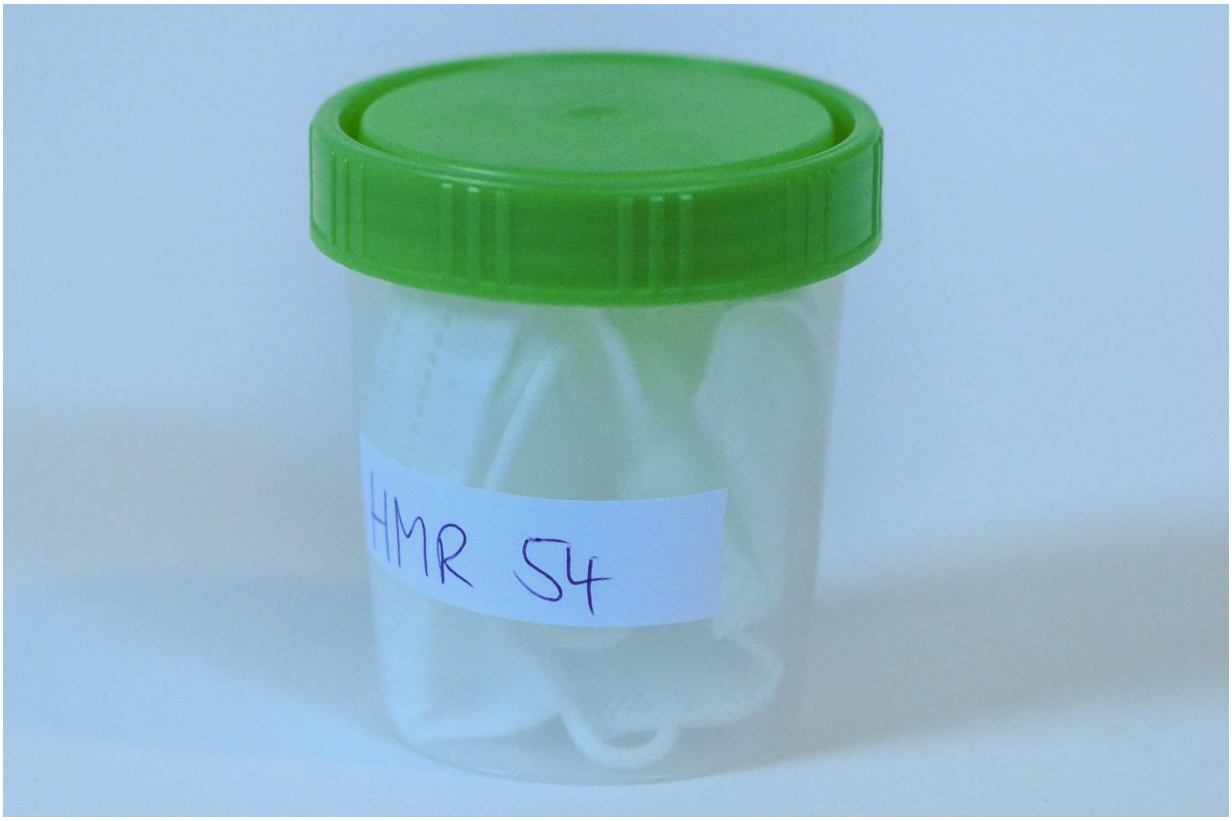

Figure 3. Stored earloop masks. After breath sample collection earloop masks were stored in plastic cups and filled in zipper-bags.

4.1.3 Dog teams, training and conditioning of dogs

Two dog teams participated in this part of the study. The two groups were trained differently, but all groups conditioned their dogs by a combination of classical and operant conditioning. Although no ethical vote is necessary for dog training and -work, all dog relevant processes were designed by experienced veterinarians or professional dog trainers.

4.1.3.1 Team 1

The dog trainer had a more than 20 years experience as police dog trainer, including many years of experience with educating tumor sniffer dogs. He worked with 5 experienced working dogs and had previously used glass tubes containing charcoal as carrier-material for cancer sample detection. Dogs were trained in a non-air-conditioned room at room temperature. The room was cleaned every day and after each trial the sniffing stations were disinfected. The sniffing stations were fixed about 40 cm above the floor and placed at a distance of about 40 cm from each other to permit easier sniffing by the dogs.

Table 2. Dog characteristics – Team 1

Methodological approach	Key Data Dogs (Name, Breed, Gender, Age)	Prior working experience
Team 1	Aimy, Golden Retriever, female, 4 years	Tracking, tumor sniffing
	Alf, German Shepherd, male, 8 years	Police dog, tumor sniffing
	Carlos, Crossbreed Dog, male, 5 years	Tumor sniffing
	Lucy, Crossbreed Dog, female, 3 years	Tumor sniffing
	Rocky, Crossbreed Dog, male, 4 years	Tumor sniffing

In a first step, the leashed dogs were guided along sample stations which were prepared with odorants smelling like food to stimulate dogs' interest to sniff at the sample stations. In the next step, the dogs were conditioned to identify cancer smell - sample stations were filled with breath samples of either LC patients or 'blank' (not ventilated) charcoal in glass tubes and the dogs encouraged to sniff them. To indicate the cancer smell the dogs were given a sit command while sniffing on such samples, followed by a positive reinforcement with food. Step by step, those 'blank' samples were replaced with breath samples from healthy subjects or patients with non-malignant lung diseases. In the last step, the samples in the setting were blinded for

the dog handler - the handler was informed of the result immediately by the trainer only after the dog's indication, and which also allowed for correction or reinforcement of dog's result. The training was completed when the dogs indicated the cancer-samples with a hit rate of more than 90%.

Table 3. Steps in Training – Team 1

Steps in training/ trial	Samples used	Task for dogs[1]	Reward	Blinding
I	· Plastic cups/ tubes filled with food-smelling odorants · No cancer-samples	· Walk along sample-stations · Sniff on each sample	No rewards	No blinding for neither handler nor trainer
II	· 5-8 cancer-samples · 17-20 blank-samples	Sit-command on cancer-samples	Food-reward for sitting in front of cancer-samples	No blinding for neither handler nor trainer
III	· 5-8 cancer-samples · 17-20 blank- and control-samples	Differentiate cancer- and control-samples	Food-reward for sitting in front of cancer-samples	No blinding for neither handler nor trainer
IV	· Random number of cancer-samples · Random number of control-samples	Sit in front of cancer-samples without command	Food-reward for indication of cancer-samples	Blinded location and identity of samples for handler, trainer unblinded
Double-blinded trials	· 12-16 blinded-samples · 1 unblinded cancer-sample · 8-12 unblinded blank- and control-samples	· Indicate blinded cancer-samples · Indicate unblinded cancer-sample	· Food-reward for every indication on blinded-samples · Food-reward for indication of unblinded cancer-sample	Blinded identity of samples for handler and trainer, blinded location for handler

[1] only new tasks for every step in training listed; Blank-sample: breath-sample carriers without breath of patients or probands; Control sample: samples of patients without cancer or healthy probands; Blinded: information is masked/ unknown; Unblinded: information is given; Excluding samples: no (positive) indication

4.1.3.2 Team 2

Two trainers were involved in dog training. One has a 16 years' experience in dog training, however, no experience with educating and training of tumor sniffer dogs. The other has over 30 years' experience as a police drug detection dog trainer and many years' experience as trainer for tumor sniffer dogs. He assisted the dogs' selection before the training started and joined the group in the last 2 months of the blinded trial phase. Team 2 worked with fleece earloop masks as breath sample carrier. Dog training took place in two rooms at room temperature. In one room a carousel was placed carrying five stations, where cups (containing earloop masks) could be fixed about 20 cm above the floor to enable easy access to samples for dog sniffing.

Figure 4. Carousel for dog testing. Dogs of Team 2 used the carousel that carries 5 stations, where cups (containing earloop masks) are fixed about 20 cm above the floor.

Another room (the 'playground') was provided with hiding places made of different materials, where breath samples, toys or food could be hidden. After each trial the rooms were aired and wiped once a week. Before the start of the methodological part 5 family dogs were selected, that had not worked as detector dogs before. The selection was based on their willingness to learn new things and their capability to be motivated.

Table 4. Dog characteristics – Team 2

Methodological approach	Key Data Dogs (Name, Breed, Gender, Age)	Prior working experience
Team 2	Emily, Flat Coated Retriever, female, 2 years	Household dog
	Gustav, Border Collie, male, 3 years	Household dog
	Kyra, Beagle, female, 1 year	Household dog
	Poldi, Labrador Retriever, male, 1 year	Household dog
	Stella, Crossbreed Dog, female, 3 years	Household dog

In a first step, dogs had to learn to walk (off-leash) along the carousel sniffing stations that carried cups filled with food. Later on, 1 (and later up to 2) cancer breath sample was placed in the carousel and the dogs were trained to indicate the cancer smell by sitting or lying in front of the sample, followed by a positive reinforcement with food or a toy. Step by step, the 4 empty stations were filled with blank samples first, then breath samples from healthy subjects and patients with non-malignant lung diseases. The setting was unblinded for the trainer, except for a few training trials short before the double blinded phase started. The training was completed when the dogs indicated cancer samples with a hit-rate of 90%. During the last 3 months of the blinded part of the study, some of the dogs showed a decreasing concentration, resulting in uncertain indications. To improve the situation the training strategy for these dogs was changed. The dogs were trained at a different place, a 'playground', to create a different and more playful working environment. Breath samples were hidden among toys and different materials (such as earloop masks prepared with food) and dogs were asked to only indicate LC samples.

Table 5. Steps in Training – Team 2

Steps in training/ trial	Samples used	Task for dogs[1]	Reward	Blinding
I	· Sample stations filled with food	· Walk along sample-stations · Sniff on stations	Food reward for sniffing on stations	No blinding for neither handler nor trainer
II	One cancer-sample within 4 empty stations	· Sniff on cancer-sample · Indicate cancer-sample by laying down or sitting in front of sample	Food-reward for sniffing on and indicating cancer-sample	No blinding for neither handler nor trainer
III	· 1-2 cancer-sample · 4 blank- or later on control-samples	· Sniff on several samples · Differentiate cancer- from blank- or control-samples	Food-reward for indicating cancer-samples	· At the beginning no blinding · Later on several blinded trials for handler and trainer
Double-blinded trials	· One blinded-sample · None or one unblinded cancer-sample · Random unblinded blank- and control-samples	· Indicate blinded cancer-samples · Indicate unblinded cancer-sample	Food-reward for indicating cancer-samples	Blinded identity of double-blinded samples for handler and trainer
IV (during double-blinded trials)	· One blinded-sample · None or one unblinded cancer-sample · Random unblinded blank- and control-samples	· Indicate blinded and unblinded cancer-samples · Exclude control-samples and hidden food/toys	Food-reward for indicating cancer-samples	Blinded identity of double-blinded samples for handler and trainer

[1] only new tasks for every step in training listed; Blank-sample: breath-sample carriers without breath of patients or probands; Control sample: samples of patients without cancer or healthy probands; Blinded: information is masked/ unknown; Unblinded: information is given; Excluding samples: no (positive) indication

4.2 Main part of the study

After the methodological approach, only one dog team was trained with improved conditioning strategies. For breath sampling fleece carrier materials in glass tubes, which had been already successfully used by Ehmann et al. [24] were employed. For the purpose of validation and comparison by eNose, breath samples were collected in parallel to those used with the dogs, thus employing two different fleece-based carrier materials, as well as direct breath testing with disposable collection bags.

4.2.1 Patients and healthy subject

Again, the study protocol was approved by the local ethics committee and all participants gave their written informed consent. We included 9 patients with LC, either SCLC (n=2) or NSCLC (n=7), with the specific histology of adenocarcinoma (n=4), squamous cell carcinoma (n=2), or other (n=1). Patients with COPD (n=22) without hints for malignant disease and healthy individuals (n=13) served as controls. Only patients with LC as primary tumor were included and COPD patients served as non-malignant controls only in order to simplify analysis of confounders. Factors that could influence the VOC composition of exhaled breath were collected. 30 male and 14 female patients and healthy subjects were enrolled in the trial (the male:female ratio was 2.1:1). The average age of subjects respectively was; cancer patients - 63 years (range 45-80), COPD patients - 65 years (range 49-79); and healthy subjects - 55 years (range 47-66). All patients and healthy subjects were recruited from the University Hospital Marburg. Breath sampling was performed in rooms of the University Hospital Marburg to assure a comparable ‘background smell’.

Table 6. Patients' characteristics validation part

Diagnosis	Subtype	Gender [n] (m/f)[1]	Age [Ø y][1] (range)	Stage of disease[2] [n]				
Validation Part				I	II	III	IV	n.a.
Lung cancer (n=9)		7/2	63 (45-80)	1	1	3	4	
	SCLC[3]	1/1					2	
	NSCLC[4]							
	Adenocarcinoma	3/1		1		2	1	
	Squamous	2/0			1	1		
	Other	1/0					1	
Non malignant lung disease (n=22)	COPD[5]	17/5	65 (49-79)	3	8	4	2	5
Healthy subjects (n=13)		6/7	55 (47-66)					

[1] m male; f female; y years; [2] based on the TNM-classification for Lung cancer; GOLD criteria for classifying stages of COPD; [3] Small-cell lung cancer (SCLC); [4] Non small-cell lung cancer (NSCLC); [5] Chronic obstructive pulmonary disease (COPD)

4.2.2 Breath sample collection

For breath sampling, fleece in glass tubes was employed as carrier material for dogs. This was done to identify differences between fleece earloop masks and fleece in glass tubes. Similar to the procedure used by Ehmann et al. [23], breath samples were collected in glass tubes containing two different (hydrophobical and hydrophilic) fleece materials (Asota, Austria), either with or without siliconisation (CHT R. Beitlich GmbH Tübingen). All fleece-containing glass tubes were prepared by the same person following a standardized procedure. For breath sampling, participants had to inspire deeply, then expire completely through the glass tubes. This was repeated five times. The tubes were closed with silicone-caps, placed in zipper-bags, sent to the dog team or used for eNose analysis. Breath sample collection with earloop masks was performed as described in chapter 4.1.2, with one mask being used for dogs and one for the eNose Cyranose 320.

Figure 5. Sample collection with fleeces in glass tubes. Breath sample collection with glass tubes containing two different (hydrophobical and hydrophilic) fleece materials. Participants had to inhale deeply then exhale completely through the glass tubes and to repeat this procedure five times.

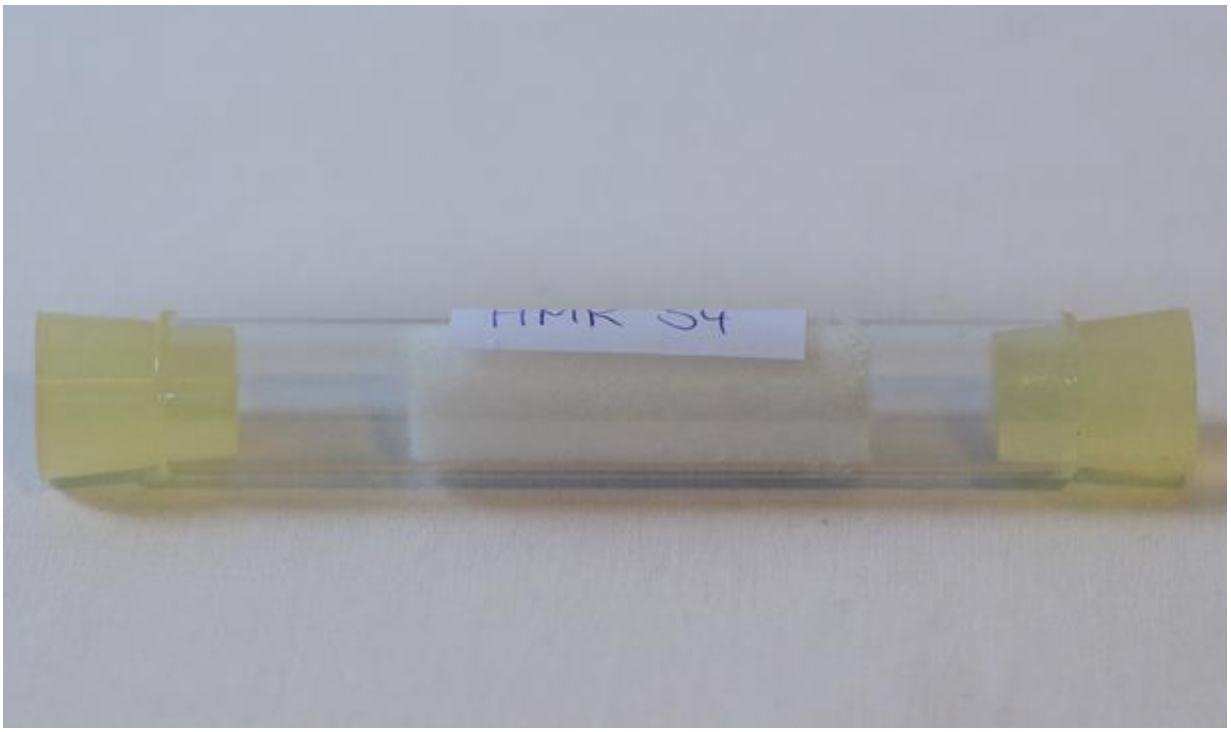

Figure 6. Stored fleeces in glass tubes. After breath sample collection the glass tubes were closed with two silicone caps and then filled in zipper-bags.

4.2.3 Dog team, training and conditioning of dogs

The main part of the study included only one dog team. All dog handlers were experienced in dog training having either worked with Search and Rescue Dogs, or as trainers for Sniffer Dog Sport or Obedience classes, but without experience in training dogs for cancer detection. There was no external trainer, but every handler was a trainer for the other dogs but not their own dog. At frequent intervals a retired police dog trainer supported the training. Training took place in a non-air-conditioned room at room temperature with three sniffing stations and in each station up to 13 samples could be presented to the dogs.

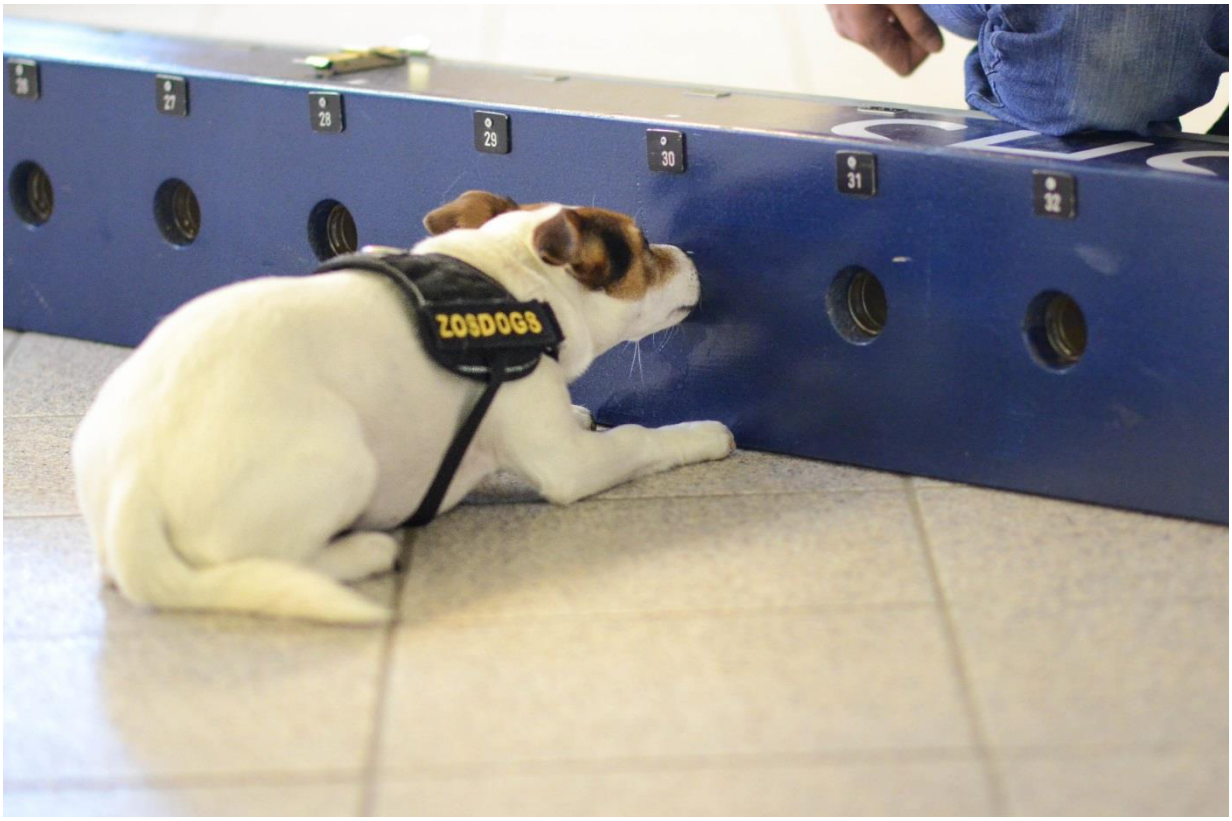

Figure 7. Sniffing station validation part. Dogs in the main used three sniffing stations for sample analysis, whereby in each station up to 13 samples can be presented.

The stations could be set on the floor or be extended so that all dogs reached the samples easily. The room and the sniffing stations were cleaned after each training day.

Dogs were used to being trained at least twice a week and were experienced sniffer sport dogs before the study started. They were educated to search for hidden targets

with only very few scents (like coins) and indicate them in a field or in sample stations comparable to those used during the study.

Table 7. Dog characteristics validation part

Validation part	Key Data Dogs (Name, Breed, Gender, Age)	Prior working experience
	Ronja, Jack Russell Terrier, female, 7 years	Sniffer Dog Sport
	Buffy, Crossbreed Dog, female, 7 years	Sniffer Dog Sport
	Mylo, Crossbreed Dog, male, 8 years	Sniffer Dog Sport
	Bizzy, Gigant Schnauzer, male, 8 years	Sniffer Dog Sport

In training, the dogs were introduced to the sampling material and to cancer breath samples first. Empty glass tubes and fleece material as well as breath samples from cancer patients were placed in the sniffing stations. The dogs learned to indicate only cancer samples by lying in front of them. Furthermore, in a second step, they learned to walk along the stations without indication when there were no cancer breath samples available (following here called 'exclusion'). Dogs were also encouraged with food after each walk, independent from an indication or exclusion. In a third step the dogs were trained to differentiate cancer samples from control samples of COPD patients or healthy subjects. Throughout all training procedures, samples were set on different positions. Dog handlers worked under blinded conditions already through the whole training period and the trainers who stayed in the room were blinded during the third step to facilitate double blind conditions. The training was completed when the dogs showed a hit rate of >90%.

Table 8. Steps in Training – Validation part

Steps in training/ trial	Samples used	Task for dogs[1]	Reward	Blinding
I	· At least one cancer sample · Blank samples put in its components (fleeces, glass tubes, silicone caps)	· Sniff on each sample · Indicate cancer samples by laying down in front	Food reward for indicating cancer samples by laying down in front	Blinded location of samples for handler and trainer, identity of samples for handler blinded, trainer unblinded
II	· None or several cancer samples · Several blank samples	· Indicate cancer samples · If no cancer samples contained no indication	· Food reward for indicating cancer samples · Food reward for excluding blank samples	Blinded location and identity of samples for handler and trainer
III	· None or several cancer samples · Several blank and control samples	Differentiate cancer and control samples	· Food reward for indicating cancer samples · Food reward for excluding blank and control samples	· Blinded location and identity of samples for handler and trainer
Double-blinded trials	· One blinded sample · One or no unblinded cancer sample · 10-11 unblinded blank and control samples	· Indicate blinded Cancer samples · Indicate unblinded cancer sample · Exclude blinded and unblinded control samples	· Food reward for every indication on (blinded and unblinded) cancer samples · Food reward for excluding control samples	Blinded identity of samples for handler and trainer, blinded location for handler

[1] only new tasks for every step in training listet; Blank-sample: breath-sample carriers without breath of patients or probands; Control sample: samples of patients without cancer or healthy probands; Blinded: information is masked/ unknown; Unblinded: information is given; Excluding samples: no (positive) indication

4.2.4 Experimental set up of the eNose

Analysis of breath samples (either direct testing with collection bags or breath samples collected with earloop masks and glass tubes filled with different fleeces) were performed in parallel to dog sniffing using the Cyranose 320 (Smiths Detection Group Ltd. Watford, UK).

Figure 8. The electronic nose Cyranose 320.

The aim was to differentiate patients with LC, those with non-malignant but inflammatory lung disease (COPD of GOLD stage 1-4), and healthy age-matched control subjects, and also to compare the results obtained with different carrier materials. Breath sample collection and measuring procedures were performed in analogy to Greulich et al., Koczulla et al. [27-28]. Patients and healthy subjects had to follow a special procedure as described before: They had to be fasting for at least two hours, including no smoking, had to flush their mouth and throat with water, and clean their nose before testing. The participants inhaled reference air (Linde) via a demand valve and exhaled with a constant flow of 100 ml/sec into a commercially available plastic bag (Rossmann). Out of the plastic bag, the eNose measured the VOCs of the exhaled breath sample. The device used contains 32 different polymeric sensors on a

nanocomposite array. The VOCs bind competitively to the sensors which triggers a change in electrical resistance at each sensor depending on size, charge, hydrogene binding capacity etc. Consequently, this generates a pattern, dependent on the chemical composition of the VOC mixture (finally it is an electric signal).

A measurement with the Cyranose 320TM consisted of three steps. 1. Baseline: Sensors were exposed to medicinal reference air. 2. Sampling: Sensors were exposed to sample air. 3. Purging: Sensors were refreshed by exposing them to ambient air. The measurements were performed in triplicate, and average values were used for calculations. Quality was checked by performing frequent measurements of room air. Also, sensor maintenance was performed repeatedly by sending the eNose to the manufacturer. Breath samples collected with different fleece materials were measured shortly after collection.

To study reproducibility of results, measurements were performed for stability analysis of VOC sampled with the fleece-based sampling materials. Therefore 4 breath samples were collected from the same patient. One was assessed right away, the other 3 samples were stored in a dark temperature controlled room and assessed after 1, 3 and 6 months.

4.3 Data analysis

For data description, mean values, standard deviations and ranges were computed. Statistical analysis were performed using Excel and SPSS software to calculate sensitivity, specificity, and positive and negative predictive values. Furthermore, chi-square and Fisher's exact tests (one/two sided) were used to evaluate statistical significance of contingency tables. In addition, binary logistic regression analysis was performed to account for confounders. Statistical significance was defined as 'highly significantly different' in cases with p-values < 0.005, as "significant" in cases with p-values between 0.005-0.05, and as 'a tendency towards significance' in cases with p-values between 0.05-0.10.

The analysis of data obtained via eNose was performed by standard multivariate bioinformatics tools. All three sample types (direct human breath, carrier materials fleece masks, fleeces in glass tubes) were compared in pairs by linear discriminant

analysis (LDA) with the aim to separate groups. The predicted values derived from the respective discriminant functions were used to calculate specificity and sensitivity. Cross-validation was performed by a k-fold cross-validation using a leave-two-out approach (one sample of each group), whereby k was equal to the product of the sample sizes of the compared groups. These analysis were performed using R software version 3.3.1.

5. Results

Parts of the results are already published in Biehl et al., Acta oncologica 2019 [1].

5.1 Methodological part

The aim of the methodological part was to yield proof of principle, that dogs can be successfully educated to discriminate breath samples of cancer patients from those of healthy subjects and patients with non-malignant lung diseases. In particular, to refine the total strategy with regard to optimized carrier material and standardized dog selection and training, it was planned to evaluate; the influence of the dogs' prior working experience; different dog training methods for the accuracy of dogs' results; and the influence of different carrier materials for breath sampling.

Sensitivity in the methodological part was 36.1% to 50%, specificity was 59.6% to 69.2%. The results of the methodological approach suggested that while specially trained dogs can differentiate breath samples of cancer patients from those of control subjects, they do so with insufficient reliability. They also suggested a major influence of the dog training methods and carrier materials. It turned out, however, that charcoal in glass tubes did not qualify as breath sample carrier material, even if the dog trainer obtained very positive results with this carrier material in the past.

5.1.1 Results of Team 1

In the double blinded phase, Team 1 captured 157 breath samples, including 68 cancer samples, and 89 control samples from non-cancer patients and healthy subjects. The collective result of five dogs regarding sensitivity indicated that 34 of 68 samples from donors with lung cancer were classified correctly (50%). Concerning specificity dogs classified 53 of 89 samples from donors without lung cancer correctly (59.6%). Concerning positive predictivity results showed, that 34 of 70 indications as cancer sample by dogs were correct (48.6%), concerning negative predictivity results showed that 53 of 87 exclusions of non-cancer samples were correct (60.9%). These differences were significant using Chi-square test (Fishers one-sided test p= 0.15; two-sided p=0.26; Chi-square test p=0.05). No results were available for the individual dogs, only collective results.

Table 9. Dogs' results – Team 1

	Breath samples with LC correct/ false (% correct[1])	Breath samples without LC correct/ false (%correct[2])	PPV[3]	NPV[4]	P[5]
Team 1	34/68 (50.0)	53/36 (59.6)	48.6	60.9	>0.1/ 0.054

[1] Sensitivity; [2] Specificity; [3] Positive predictive value; [4] Negative predictive value; [5] P-values obtained with Fisher's exact test, one sided/ Chi-square test

Some breath samples had been taken twice and were offered later a second time to the dogs in order to study reproducibility of dogs' indication. Reproducibility analysis by Team 1 were performed with 33 breath-samples; 9 LC-samples, 13 samples from patients with non-malignant lung diseases and 13 samples from healthy probands using the same dogs as before. In 15 of 33 cases (45.5%) the dogs did not indicate the same results (either positive or negative) as before if patients' samples were presented a second time, so reproducibility of results was insufficient.

5.1.2 Results of Team 2

In the blinded phase of the trial 88 breath samples were included, 36 cancer-samples and 52 control samples from healthy subjects and patients with non-malignant lung diseases. The collective result of five dogs for specificity indicated that 36 of 52 samples from donors without LC were recognized correctly (69.2%). Concerning sensitivity dogs indicated 13 of 36 samples from donors with LC correctly (36.1%). Concerning positive predictivity data showed, that 13 of 29 results as indicated by dogs were correct (44.8%), negative predictivity showed that 36 of 59 exclusions of non-cancer samples were correct (61%). These results were statistically not significant (Fishers exact test and Chi-square test: $p > 0.1$).

Training was changed in the last 3 months of the blinded trial for two of the dogs of Team 2 because of a decreasing concentration in order to improve their indications. Improved training with these two dogs yielded better results, although case numbers were too low to yield statistical significance. Concerning sensitivity dog Poldy indicated 10 of 19 samples correctly (52.6%), dog Gustav indicated 7 of 14 samples correctly (50%). Dog Poldy achieved a specificity of 74% (25 of 34 samples from donors without LC excluded correctly), dog Gustav achieved a specificity of 67% (14 of 21 samples from donors without LC excluded correctly). Concerning positive predictivity 10 of 22 cancer samples indicated by Poldy as cancer samples were correctly (46%), Gustav 7

of 16 samples (44%). Differences obtained by Poldy were statistically significant using the Qui-square test (Qui-square p=0.018; fishers one-sided test p= 0.12; two-sided p=0.16). So the data suggest, that Poldy (but not Gustav) could significantly differentiate non-cancer and well as cancer samples.

Table 10. Dogs‘ results – Team 2

	Breath samples with LC	Breath samples without LC	PPV[3]	NPV[4]	P[5]
	correct/ false (% correct[1])	correct/ false (% correct[2])			
Team 2	13/36 (36.1)	36/16 (69.2)	44.8	61.0	>0.1/ >0.1

[1] Sensitivity; [2] Specificity; [3] Positive predictive value; [4] Negative predictive value; [5] P-values obtained with Fisher’s exact test, one sided/ Chi-square test

Reproducibility of results of breath samples from the same participants collected twice could be tested with 18 samples: 5 cancer samples, 9 samples from patients with non-malignant lung diseases and 4 samples from healthy subjects using the same dogs as before. Results were similar to Team 1: Reproducibility of results was insufficient, in 8 of 18 (44.4%) cases the dogs did not achieve the same results (either positive or negative) as before if patients' samples were presented a second time.

5.1.3 Analysis of confounders

Data about concommitant diseases, medications as well as nutrition or consumption habits of all patients were collected. Since hit rates in the teams were unsatisfying informative value of analysis of confounders was low. Still no influences of confounders on the dogs’ results could be detected.

5.2 Main study

Based on the findings of the methodological approach an improved selection of dogs, as well as improved dog training and additional sampling material were used in the main part of the study. To exclude a potential failure of fleece earloop masks fleece in glass tubes, which have already been used successfully by previous investigators (Ehmann et al., McCulloch et al.), were tested. Furthermore the two different fleece based sampling materials were compared using an eNose for analysis and in the end eNose and dogs’ results were compared.

Dogs' results to differentiate cancer breath samples from control samples were statistical significant with a sensitivity of 55.6% and a specificity of 82.9%. Analysis of individual dogs' results showed big differences in accuracy of differentiation. Results obtained with eNose showed comparable results for direct testing and testing with fleece based breath sampling materials with sensitivities between 89-100% and specificities between 97-100%. Even if dogs were able to significantly differentiate cancer from non-cancer samples, the eNose achieved better results for both, sensitivity and specificity.

5.2.1 Results obtained by dogs

In the blinded phase of the trial 44 breath samples were included, 9 cancer-samples and 35 control samples from healthy subjects and patients with COPD. Collective results obtained as an average of the four individual dogs' results showed that 5 of 9 cancer samples were correctly indicated with a sensitivity of 55.6%. 29 exclusions of 35 non-cancer samples were correct, yielding a specificity of 82.9%. Concerning the positive predictive value, 5 of 11 indications were correct (46%), concerning the negative predictive value, 29 of 33 exclusions were correct (88%). These results correspond with a statistically significant differentiation (Chi-square p=0.018, fisher's one sided test p=0.03, two-sided p=0.03).

Individual dog's results were available for 44 breath samples for dogs Mylow, Buffy and Ronja, and for 39 breath samples for dog Bizzy. Dogs Mylow and Buffy achieved statistically significant results (Chi-square p=0.033 and 0.034; Fisher's exact test one sided p=0.047 and 0.044, two sided p=0.087 and 0.053, respectively) for all breath samples, while dogs Ronja and Bizzy did not achieve significant results (Chi-square p=0.40 and 0.089; Fisher's exact test one sided p=0.356 and 0.123, two sided p=0.586 and 0.123). Mylow and Buffy achieved a specificity of 80 and 71%, while dogs Ronja and Bizzy achieved values of 89 and 82%. Results for sensitivity for Mylow and Buffy were 56 and 67%, while Ronja and Bizzy achieved a sensitivity of 22 and 50%. Concerning the positive predictive value, Mylow and Buffy achieved 42 and 38%, and Ronja and Bizzy 33 and 33%. The negative predictive value shown by Mylow and Buffy was 88 and 89%, and that for Ronja and Bizzy was 82 and 90%. These data show that Mylow and Buffy could differentiate non-tumor from tumor samples, while Ronja and Bizzy could not.

Table 11. Dogs' Results Validation part

	Breath samples with LC correct/ false (% correct[1])	Breath samples without LC correct/ false (%correct[2])	PPV[3]	NPV[4]	P[5]
Collective Results	5/9 (55.6)	29/6 (82.9)	45.5	87.9	0.03/ 0.018
Mylow	5/4 (55.6)	28/7 (80.0)	41.7	87.5	0.047/ 0.033
Buffy	6/3 (66.7)	25/10 (71.4)	37.5	89.3	0.044/ 0.034
Ronja	2/7 (22.2)	31/4 (88.6)	33.3	81.6	>0.1/ >0.1
Bizzi	3/3 (50.0)	27/6 (81.8)	33.3	90.0	>0.1/ 0.089

[1] Sensitivity; [2] Specificity; [3] Positive predictive value in %; [4] Negative predictive value in %; [5] P-values obtained with Fisher's exact test, one sided/ Chi-square test

Determination whether the presence of concomitant diseases, smoking habits or other characteristics of the patients influenced the results of sniffing was done. The presence of LC was correlated with the Body Mass Index (BMI, p=0.004), whereby LC patients were characterized by lower BMI values. Furthermore, there was a relationship of LC to current smoking habits (Fisher's exact test, one-sided p=0.004; chi-square p=0.001), but no correlation of LC with COPD could be shown (Fisher's exact test one-sided p=0.435; Chi-square p=0.582). There was no significant correlation of LC with the coexistence of cardiovascular disease and/or diabetes mellitus (Fisher's exact test one-sided p=0.298; Chi-square p=0.362) but a correlation with the presence of gastrointestinal diseases (Fisher's exact test one-sided p=0.001; Chi-square p< 0.001). Conversely, there was no correlation between gastrointestinal disease and COPD (Fisher's exact test one-sided p=0.579; chi-square p=0.687).

Table 12. Participants' concomitant diseases – validation part

	LC[1] patients (n=9)		COPD[2] patients (n=22)		Healthy subjects (n=13)	
Concomitant disease	n	%	n	%	n	%
Cardiovascular disease	4	44.4	6	27.3	3	23.1
Diabetes mellitus	1	11.1	3	13.6	0	0.0
Thyroid disease	2	22.2	1	4.6	1	7.7
Gastrointestinal disease	4	44.4	0	0.0	0	0.0
Renal disease	1	11.1	1	4.6	0	0.0
Other	5	55.6	7	31.8	1	7.7
None	0	0.0	10	45.5	9	69.2

[1] Lung-cancer; [2] Chronic obstructive pneumological disease

The data revealed some associations between factors potentially influencing the sniffing results in the group of study subjects, which raised the question of a potential influence on the results of the dogs' indications. Logistic regression analysis using both age and BMI as predictors did not reveal a relationship of age (p=0.315) or BMI (p=0.945) to the dogs' indications, suggesting that the indications were not influenced by these two factors. The indications were also not related to the presence of COPD within a 2x2 contingency table (chi-square p=0.101). This was in line with the results of logistic regression analysis using both the presence of LC and COPD as predictors of indications and demonstrating a significant relationship of indications to LC (p=0.014) but not to COPD (p=0.135). The dogs' indications were also independent from the patients' smoking habits (Fisher's exact test one-sided p=0.267; chi-square p=0.315). Despite the correlation of gastrointestinal disease with the presence of LC, the dogs' indications were not influenced by the presence of gastrointestinal disease (Fisher's exact test one-sided p=0.256; chi-square p=0.226). No correlation between LC and inhalative medications of LC patients (Fisher's exact test one-sided p=0.320; chi-square p=0.401) was seen, but there was a clear correlation with systemic medications of LC patients (Fisher's exact test one-sided p=0.007; chi-square p=0.008), which is understandable, as LC patients regularly receive systemic medication.

Within the limited data set, no multiple covariates beyond one confounder could be reliably tested, but the various analysis using one covariate in addition to LC as predictor did not reveal other relevant influencing factors on the dogs' indications than LC, especially COPD had no influence. Only 2 of 9 LC patients were in an early stage of the disease (stage 1 to 2), whereas 7 patients suffered from advanced tumor stages

(stage 3 to 4). Both samples from the early stage LC patients were indicated as negative by the dogs, pointing towards a correlation of low tumor burden with false negative indications.

5.2.2 Results obtained with the eNose

In addition to the breath samples presented to the dogs, for all participants parallel samples were collected for eNose analysis. This was achieved in three ways, either by direct breath sampling using a collection bag, by breath sampling with earloop masks followed by later eNose analysis, or by glass tubes filled with different fleeces followed by later eNose analysis.

Table 13. Results obtained with eNose

Breath sample collection type	Breath samples with LC correct/ false (% correct[1])	Breath samples without LC correct/ false (% correct[2])	PPV[3]	NPV[4]	CCV[5]
Direct testing	9/0 (100)	35/0 (100)	100	100	52.2
Earloop masks	9/0 (100)	34/1 (97.1)	90	100	35.4
Fleece	8/1 (88.9)	30/1 (96.8)	88.9	96.8	42.4

[1] Sensitivity; [2] Specificity; [3] Positive predictive value in %; [4] Negative predictive value in %;
[5] Cross- validation- value in %

The eNose analysis of samples that were directly collected in bags in the 44 participants (9 LC, 22 COPD, 13 healthy) yielded a 100% correct differentiation between samples, corresponding to 100% values for sensitivity and specificity in all comparisons. The cross-validation value (CVV) for differentiation of LC from the group of other participants including COPD patients and healthy subjects was 52%. The respective CVV for the differentiation of COPD patients without LC from healthy subjects was 49%, and that for the differentiation of COPD patients irrespective of the presence of LC from healthy subjects was 49%; in this pooled comparison, sensitivity was 97% and specificity 92%.

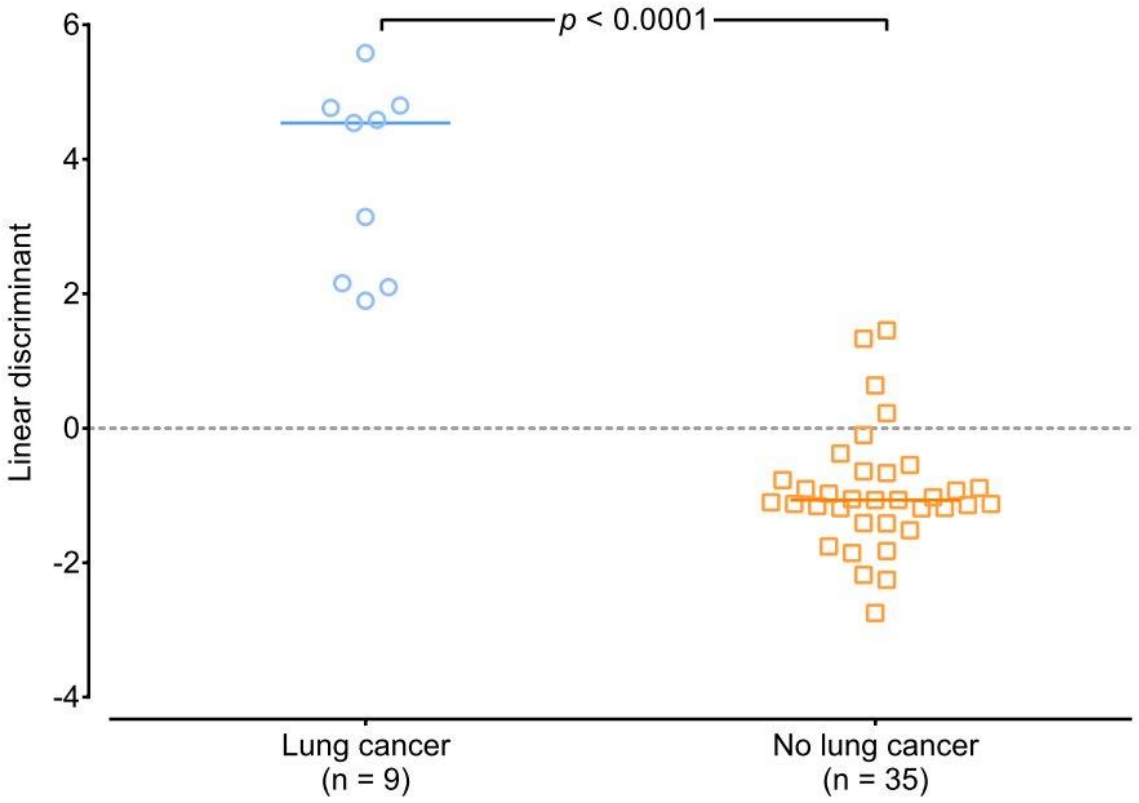

Figure 9. Comparison of LC and non-cancer samples – direct testing. VOC differentiation after direct testing of breath samples collected in collection bags by eNose between patients with LC (circles) or without LC (squares). The vertical axes show the values of the pairwise discriminant function based on the results of all 32 sensor signals.

The analysis of fleece earloop masks as carriers (stored until use) via eNose was performed for 44 participants (9 LC, 22 COPD, 13 healthy) and showed a high capability of differentiation between samples. Regarding specificity, 34 of 35 (97%) of samples from donors without LC were correctly recognized. Regarding sensitivity, the eNose indicated 9 of 9 (100%) of LC samples correctly. Regarding positive predictive value, data showed that 9 of 10 positive results obtained by eNose were correct (90%). Regarding negative predictive value, 34 of 34 negative results were correct (100%). The CVV for differentiation of LC from the control group (COPD plus healthy subjects) was 35%. Differentiation of COPD without LC from healthy subjects showed a sensitivity and specificity of 100% (CVV 74%), differentiation of all COPD patients, including those with LC, from healthy subjects also achieved a sensitivity and specificity of 100% (CVV 56%).

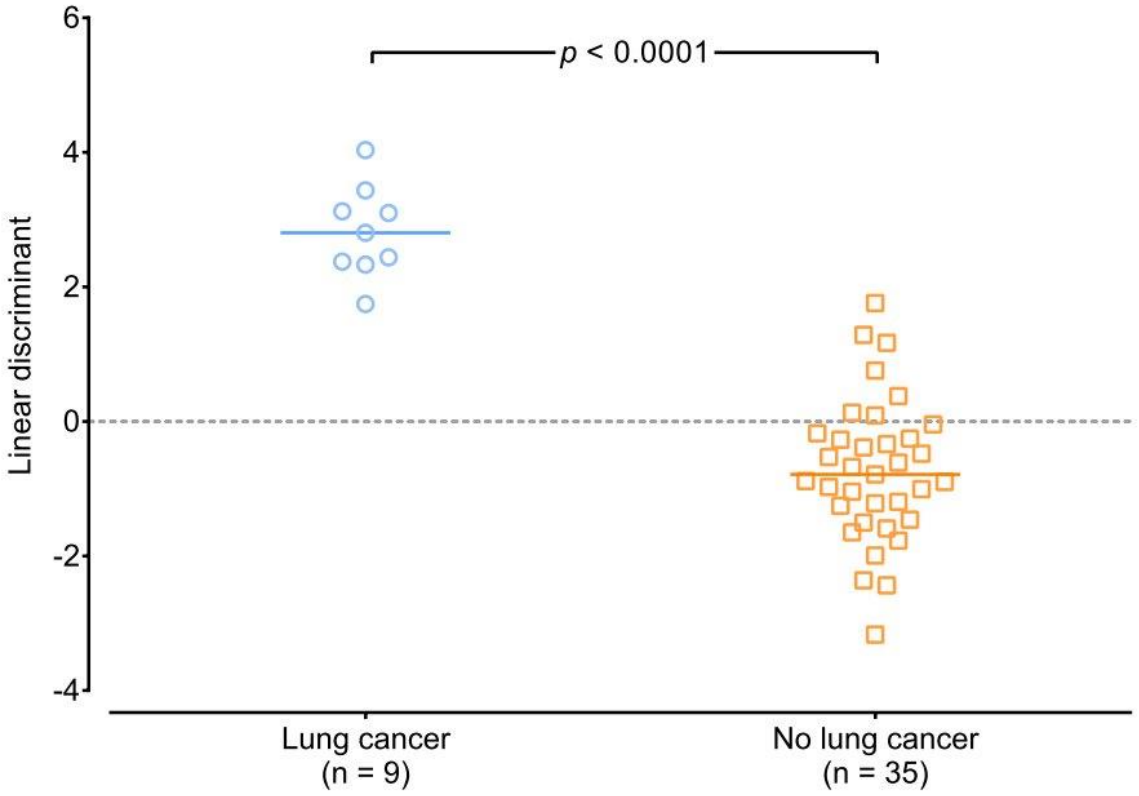

Figure 10. Comparison of LC and non-cancer samples – earloop masks. VOC differentiation after breath analysis of earloop masks by eNose between patients with LC (circles) or without LC (squares). The vertical axes show the values of the pairwise discriminant function based on the results of all 32 sensor signals.

Three COPD patients and one healthy subject could not provide breath samples with fleece filled glass tubes, due to exhausting breath sampling (COPD in end stage disease and exhausted control patient). The eNose analysis of fleece-filled glass tubes as breath samplers were thus performed in the remaining 40 patients (9 LC, 19 COPD, 12 healthy) and showed a significant discrimination between samples. Overall, 30 of 31 samples from subjects without LC were recognized correctly (specificity 97%), and 8 of 9 LC samples (sensitivity 89%). Concerning the positive predictive value, data showed that 8 of 9 results were correct (89%). Concerning the negative predictive value, 30 of 31 exclusions were correct (97%). The CVV for differentiation of LC from the pooled control group was 43%. Furthermore, the eNose differentiated COPD patients without LC from healthy subjects with an overall sensitivity of 95% and specificity of 100% (CVV 39%). In the differentiation of all COPD patients, including those with LC, from healthy subjects, the eNose achieved a sensitivity of 96% and specificity of 83% (CVV 38%). In summary, the data showed, that the eNose yielded comparable results, when using either direct testing or carrier materials for breath

sampling, and was able to differentiate LC, COPD and healthy controls with a high sensitivity and specificity.

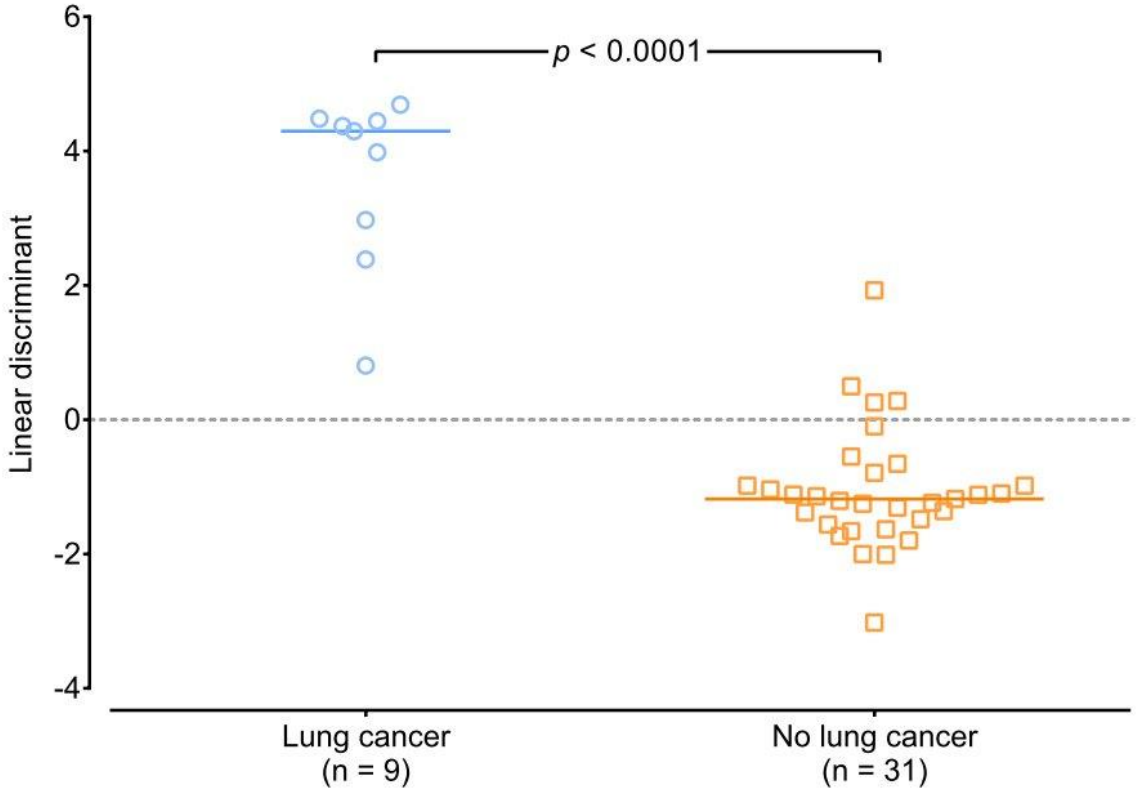

Figure 11. Comparison of LC and non-cancer samples – fleeces in glass tubes. VOC differentiation after breath analysis of glass tubes filled with different fleeces by eNose between patients with LC (circles) or without LC (squares). The vertical axes show the values of the pairwise discriminant function based on the results of all 32 sensor signals.

Concerning analysis of confounders, a specific confounder that influenced the eNose results could not be identified, so there were no suggestions that the presence of concomitant diseases, smoking habits or other patient-associated characteristics, was important. In the differentiation between LC samples and non-cancer samples, one breath sample was indicated as false positive (healthy subject, male) and one breath-sample as false negative (LC patient with adenocarcinoma stage IV, male). In the differentiation between COPD samples (with and without LC) and samples of healthy subjects, three samples were indicated as false positive (healthy subject, female, n=2; healthy subject, male, n=1), and two breath samples were indicated as false negative (COPD patient, male; COPD patient, female).

5.2.2.1 Stability analysis of VOC sampled via earloop masks and glass tubes

Breath sampling with fleece based sampling material as earloop masks and fleece in glass tubes provided a feasible way to collect VOC from breath samples. The samples could be stored and did not have to be analyzed with the eNose directly. The samples of 3 patients were studied at four different time points (day 0, 1 month, 3 months, and 6 months later) and the results compared. Most of the 32 sensors responded in a stable and reproducible way over time. However, VOC sensors 6, 23, 28, and especially 31, showed a decrease of their signal for both fleece earloop masks and fleece in glass tubes. Thus, most VOC appear to be quite stable on earloop masks as well as fleece in glass tubes for at least 1 month, though cannot be stored longer than 3 months.

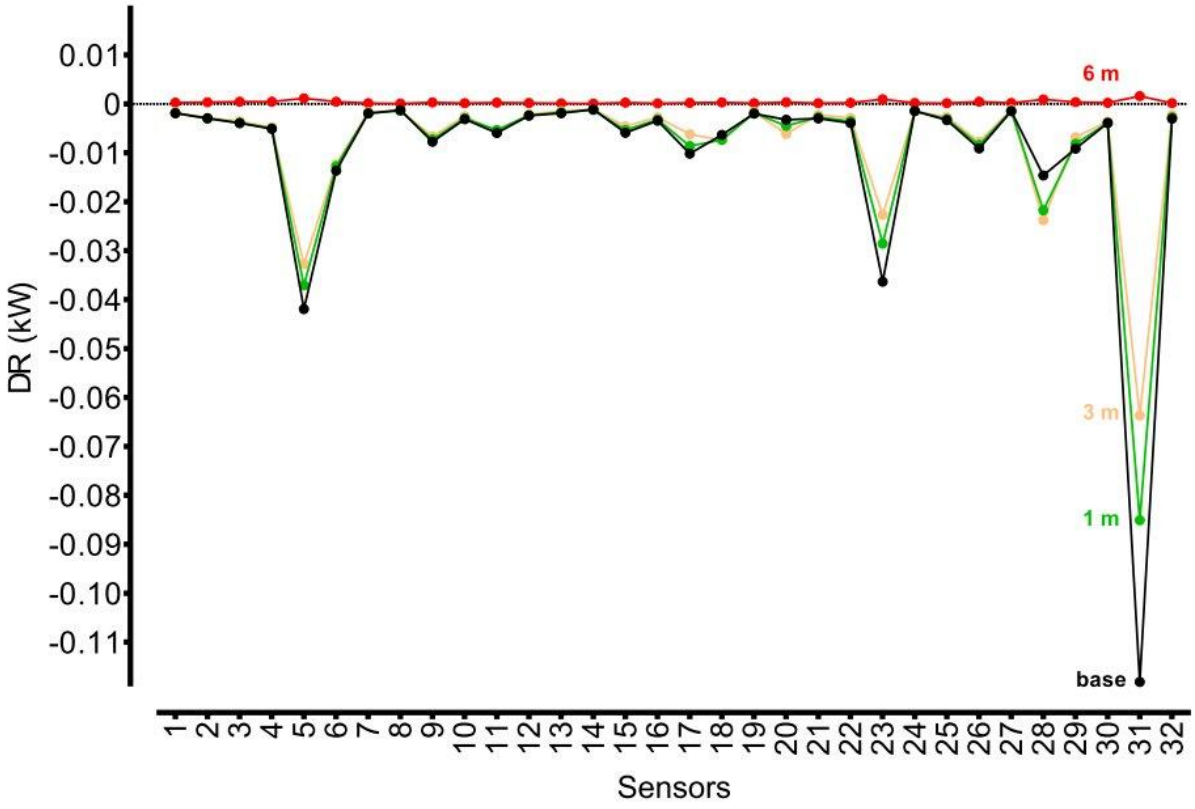

Figure 12. Stability analysis - earloop mask. Repeated VOC analyses of a fleece earloop mask sample by eNose over 6 months. The sensor responses of one fleece earloop mask sample are given. Over time periods of 1 month (1 m), 3 months (3 m), and 6 months (6 m) VOC were stable for most sensors. The responses of sensors 6, 23, 28, and 31 decreased (?R increases) over time. After 6 months the signal was no longer detectable.

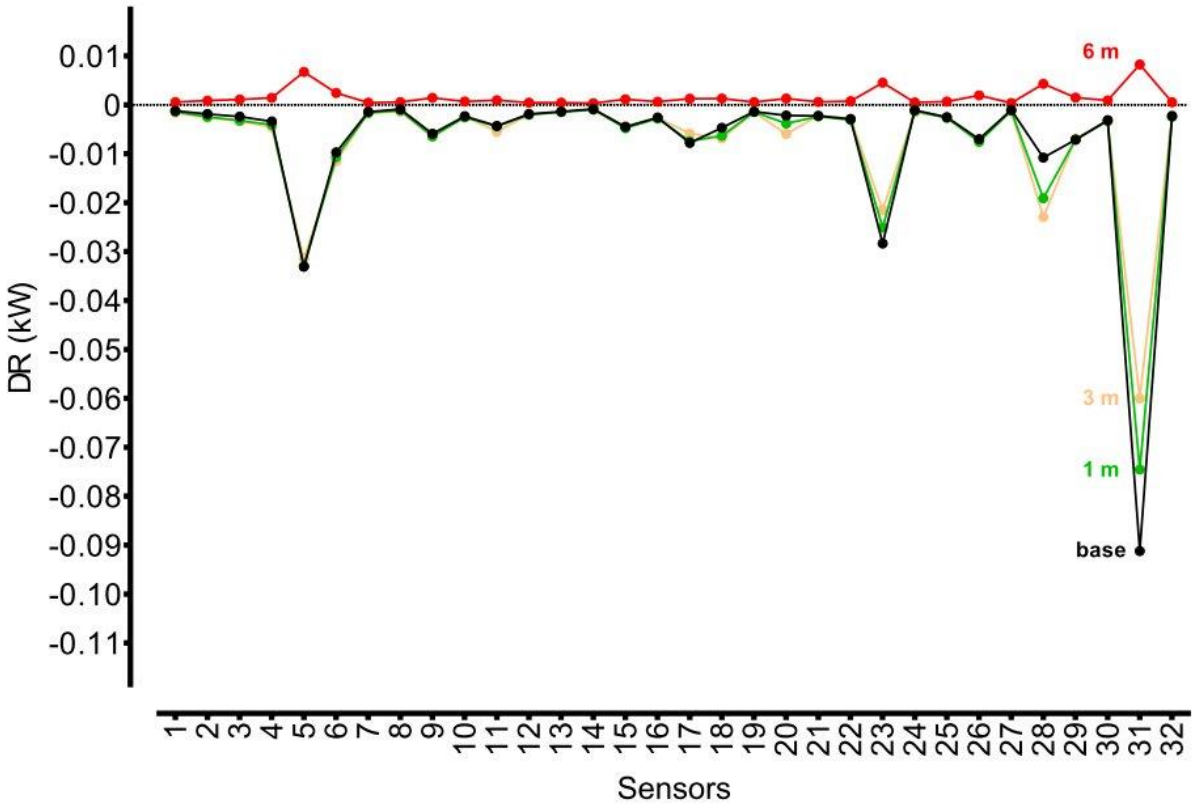

Figure 13. Stability analysis – fleeces in glass tube. Repeated VOC analyses of different fleeces in a glass tube sample by eNose over 6 months. The sensor responses of one fleece in glass tubes sample are given. Over time periods of 1 month (1 m), 3 months (3 m), and 6 months (6 m) VOC were stable for most sensors. The responses of sensors 6, 23, 28, and 31 decreased (?R increases) over time. After 6 months the signal was no longer detectable.

5.2.3 Relationship between the results obtained by dogs and eNose

To evaluate the capabilities of eNose and dogs, the results of the main study obtained for breath samples collected with fleece in glass tubes were compared. Even if dogs were able to significantly differentiate cancer from non-cancer samples, the eNose achieved better results for both, sensitivity and specificity.

Two of these 'false negative' indications by the dogs referred to patients in an early cancer stage, whereas the eNose indicated a patient in a late cancer stage as false negative. No correlations were found between the individual false positive and false negative results as obtained by the eNose and the dogs. Furthermore, no common confounders could be identified. This suggests that the strengths and failures of both approaches were based on different sample characteristics.

6. Discussion

Parts of the discussion are already published in Biehl et al., Acta oncologica 2019 [1].

6.1 Screening for lung cancer using sniffer dogs

Recently papers have been published on canine scent detection using materials containing volatile ‘tumor associated markers’ in tumor tissue, urine, watery stool samples or exhaled breath [20-26]. The studies showed that dogs are capable of identifying volatile marker profiles associated with cancer with remarkable accuracy (sensitivities 41-99%, specificities 91-99%). Furthermore, these studies suggested that dogs' indications were not influenced by concomitant diseases, medications, consumption habits or smoking. Usually the dogs had to identify one cancer breath sample within several control samples. In contrast, two other studies showed inferior results, the sensitivity of dogs being 45-76% and specificity 8-53%, if dogs were confronted with a real-life screening-like situation and when varying the number of LC samples (between 0 and 5) in the double-blind setting [24-26]. In the methodological approach, sensitivity was 36-59% and specificity 45-60%, while in the main study sensitivity was 56% and specificity 83%, underlining the importance of the experimental settings for the results. Taken together, the data confirm that dogs' noses in general have the capability of identifying cancer smell and differentiating it from 'background smell'. However, large variations in the dogs' correct indications for cancer and non-cancer samples occurred within different settings. These variations probably point to important confounders such as trainers, conditioning strategy, carrier materials, the dogs' individual capability, as well as patients’ individual or disease-associated characteristics which might influence the results regarding sensitivity and specificity.

6.1.1 Influence of dogs’ working experience, conditioning strategies and trainers

In the methodological approach results of experienced working dogs were compared with those of household dogs. The overall result was a slight advantage for the working dog team. Differences were found in the training period, which took considerably longer for the household dogs. Furthermore, with a lack of concentration and a refusal to work, some of the household dogs needed a training adjustment

during the double-blinded testing. The experienced working dogs needed a shorter training period and they achieved slightly better results even with less suitable breath sample material.

The data of the methodological approach show, that all dogs had a hit rate of more than 90% in the disclosed phase of the study, however, this rate was reduced during the blinded phase. These differences probably point to the fact that dog trainers have a very strong influence on the dogs' behavior and hit rates considering the personal relationship that exists between them and the potential for unconscious suggestion. In the disclosed parts of the study, this could have led to high proportions of 'correct' results, and in the undisclosed parts, to confusions of the dogs' work and elevated proportions of 'false' results – despite high correct hit rates in the disclosed trial part. Helton [28] describes the importance of undisclosed double-blinded settings during dog training. With knowing the right answer, the trainer unconsciously gives signs to the dog, which lead the dog to respond correctly. To exclude the trainers' behavioral influences on the results, dog studies should be performed under completely blinded conditions.

In the blinded part of the preliminary study, dogs were confronted with a situation, in which their performance was less rewarded than in the disclosed part, in this case with a setting with a lower proportion of LC samples. This could have led to higher numbers of false positive indications during the blinded phase of the preliminary trial, since dogs were only rewarded for positive LC indications. Furthermore, in case the dogs should have been mistakenly rewarded for a false positive indication during the blinded trial, they would have been conditioned with the wrong signals. It has already been shown that mistakes during the blinded trial can lead to an "unlearning/ forgetting" by the dog, and furthermore, a retroactive interference could affect their trained behavior [30]. The conditioning strategy for the dogs has to be chosen with great care. A rewarding of dogs' work has to be independent of the results achieved and should only relate to the work done. In case dogs are only rewarded for positive indications (as in the methodological part), they will quickly learn to achieve more rewards through positive indications, which could lead to higher false positive results and a loss of the tumor sniffing competence. A loss of tumor sniffing competence could also result if dogs are not regularly confronted with tumor smell samples.

In the main part of the study the dog team worked under new improved conditioning strategies. To minimise unwanted trainer influence of dogs' behavior, dogs and trainers worked under blinded conditions during the training period. Furthermore, dogs were encouraged for excluding non-malignant samples, so the rate of rewarding was always the same, independent of the number of cancer samples. During double-blind testing overall specificity increased up to 83%. These results suggest a high number of rewards for excluding non-malignant samples increases specificity. Still overall sensitivity of 56% was insufficient - possibly explained by higher numbers of control samples and lower numbers of LC samples in the second part of the study, leading to a loss of tumor smell recognition by the dogs. Similar to Hackner et al. [26] we suggest that a loss of tumor sniffing competence could be the result, if when dogs are not regularly confronted with tumor smell samples.

The reproducibility of the results was weak in the blinded phase within the methodological approach, possibly suggesting that dogs were irritated by the repeated presentation of the same patients' sample. With indicating, dogs could either show that a tumor scent is detected or that the dog positively confirmed that he recognized a person's smell that had already been presented before. However, weak reproducibility of results shows an uncertainty when dogs are used as a screening tool.

6.1.2 Individual dogs' capability to differentiate tumor scent from non-tumor scent

Literature data shows that some dog trainers included only one dog in scent detection [21], whereas others included 5-6 dogs and also collected the individual dogs' data [23-26]. Whereas McCulloch et al. [23] state the sniffing quality of all dogs was comparable and therefore the results obtained also comparable. Ehmann et al. [24] found differences in hit rates between individual dogs and therefore defined a 'collective dog decision' that required at least 3 out of 5 dogs reach the identical decision. Amundsen et al. as well as Hackner et al. [25-26] also showed considerable variations in dogs' individual results. These variations might be due to the dogs' different sniffing capabilities as well as the dogs' different daily conditions and the dog training. In the main part of the study, dogs' individual results showed great differences concerning sensitivity in the range of 22-67% and specificity in the range of 71-89%. We conclude that it is advisable not to rely on an individual dog's decision, but to define a collective decision in order to minimize the impact of variations from the

individual dogs' conclusions. Furthermore in a double-blind setting, the hit-rate of individual dogs should be control tested frequently to permit positive adaptation of dogs' training.

6.1.3 Dogs' capability to detect tumors depending on the stage of the disease

Detection of early phases of LC, for example in heavy smokers and/or patients with COPD, may permit earlier initiation of therapies in these patients. In our methodological approach, no differential examination was performed. In the main part of the study, only 2 patients with early stage tumors participated. None of the early stage tumors was detected by the dogs, which could be explained by a low tumor load or tumor scent, although no reliable conclusions can be drawn with such a small dataset of only 2 early stage tumors. Further investigations are needed to understand how tumors at different stages influence dogs' indications. However, it remains questionable whether dogs will be capable to detect early stages of tumor with sufficient reliability.

6.2 Volatile markers for screening of lung cancer detected by eNose

Recently, studies have been published~~,~~ in which exhaled breath collected with collection bags was analysed with an eNose in order to identify a specific 'breathprint' profile of LC [14-18]. Their results indicate that an eNose can differentiate LC 'breathprints' from those of control groups (patients with non-malignant lung disease and healthy subjects) with high accuracy. eNoses provide a non-invasive, easy, onsite, and not too expensive, tool for the screening of breath samples, although the lack of accepted protocols for standardization and the use of different eNoses currently complicate their clinical use.

The high detection capacities of an eNose can also lead to various VOC signals and exhaled VOC patterns. A wide range of calibrating samples for eNose could help to figure out influences of VOC patterns like VOC in the environment or vitiated materials.

Improved tumor detection may be achieved by the use of only one eNose for an expanded group of patients. This could be made possible by using (storable) breath sampling materials instead of direct testing (see also Chapter 6.3.2).

While eNose results for sensitivity and specificity showed a high accuracy, the CVVs was relatively low. This could be explained by the small dataset of samples, so the CVV is prone to results with a larger variance in the measurement.

6.3 Breath sample materials for samples collection

6.3.1 Role of carrier materials for dogs

One dog team (Team 1) in the methodological part used a plastic balloon for breath collection with consecutive binding of volatile molecules on charcoal in cylindrical glass tubes. Although, according to the trainer, the results obtained in training with disclosed diagnoses showed that dogs could differentiate healthy from LC samples, however, the blinded part showed only a tendency towards significant results. It is assumed that in using a plastic balloon for breath collection synthetic organic molecules of plastic may dilute or mask organic breath molecules and thereby impairs detection capability. However, according to the manufacturer's product description, using charcoal for breath collection in tubes, at room temperature, yields strong irreversible binding of organic molecules to the charcoal, and that these bound VOC when used in standard procedures are only released by heating or chemical treatment, probably leading to (partial) destruction of volatile molecules. Such heating procedures were not applied in the setting of dogs' examination in the methodological part of study. It might be speculated that molecules adhered to the remaining free glass walls provided the smell. Nevertheless, the dog trainer confirmed previous experience of positive results with the charcoal tubes and expressed a preference to work with these during the methodological phase.

For the main study, no plastic balloons and charcoal-filled glass tubes were used. According to the methodological phase, standard fleece earloop masks, as normally used in physicians' daily work, qualified as carrier materials. The use of fleece masks needs a careful handling both on the patients' as well as the samplers' side to avoid olfactory contaminations. The dogs might detect volatile breath molecules as well as dermal molecules attached to the masks, which could be bound to the masks by

patients' breath molecules as well as skin particles. Since malignant tumors are systemic diseases, it seems reasonable to assume that the combination of breath and dermal particles might positively influence the results.

The dogs in the main part of the study were confronted with glass tubes filled with fleece materials, which had been already successfully tested by Ehmann et al. [24] and achieved significant results. Both, fleece-filled glass tubes and fleece earloop masks, seem to qualify for breath sampling. Fleece materials, either with or without further processing, have the capability of binding VOC reversibly and of releasing the molecules which can be smelled by the dogs and detected by eNose. Considering the higher cost of glass tubes as well as the time-consuming preparation (siliconisation, fleece filling), it appears that fleece earloop masks for breath sampling of VOC are to be preferred.

6.3.2 Role of carrier materials for eNose

In the setting three different carrier materials for breath sample collection were tested. To our knowledge, no such comparison for an eNose has been done so far. In parallel comparisons, the best results for the differentiation between LC and control samples were obtained by direct breath collection via bags. However, results obtained with storable breath sample carrier materials were comparable to those obtained with direct testing. The use of fleece earloop masks as carriers yielded similar results as that of fleece in glass tubes. This implies that breath samples could be collected with these carrier materials and later be sent to the eNose laboratory for analysis in less than 3 months, after which the loss of VOC is substantive. With respect to patients with severe disease, a sample collection on the wards or in a doctor's office would facilitate sampling and in consequence enable the inclusion of more samples in a study cohort, especially of more high-risk patients. This strategy could also contribute to experiments with the same eNose, thereby reducing apparatus-related variations [16]. The variation in background smells during sample collection needs to be checked as a potential confounder in these studies if collection is performed at different places. It is possible, with mapping a wider range of breathing patterns (and potential confounders), the identification of a specific cancer-associated pattern could be made easier.

6.4 Limitations of the study

A major limitation of our study was the small number of breath samples collected, especially in the validation part of the study. For the methodological approach 158 samples were collected, but the results were not definitive, due to a high degree of heterogeneity among the settings and dog teams. This experience required us to develop a higher standardization in the main study, such that, new modified and improved strategies permitted statistically significant results. However, detailed, comprehensive analysis of potential confounders, such as tumor size, concomitant diseases and metabolic alterations, which might influence the results of dogs and eNose, were not possible, again due to the small patient sample population size, even though the small sample did not hint the confounders were impactful.

A particular limitation of working with dogs was the potential influence of dogs' daily constitution on sniffing performance. Different from technical systems, dogs present in various dispositions and health conditions, respond inconsistently to changing environmental [31] and they can forget what they have learned. This sets natural limits when considering their potential practical use. In addition, problems in training appeared, especially in the preliminary part of the study. While some teams trained their dogs two or three times a week, other teams trained once a week at most, resulting in an extended overall duration of training before the dogs were prepared for double-blind testing. These differences in dog training might have influenced the results.

7. Concluding remarks

Parts of the concluding remarks are already published in Biehl et al., Acta oncologica 2019 [1].

The use of breath samples to detect LC offers the possibility of a non-invasive screening tool. In an intensive methodological phase different dog selections and training strategies were tested and results showed that selected and specifically trained dogs (either working dogs or ordinary household dogs) are, in principle, able to differentiate healthy and non-malignant from malignant breath samples. Consequently, experienced working dogs are easier and quicker to train, and achieve better results, so the recommendation is to choose experienced working dogs for cancer detection. No correlation of results with nutrition and consumption habits, medication and concomitant disease of the patients could be found. In the second validating part, with improved selection and training of dogs, better results especially for specificity were achieved, while results for sensitivity still remain unsatisfying. Furthermore large differences appeared in the individual dogs' ability to detect and differentiate cancer smell. Although dogs showed the ability to detect and differentiate LC smell from non-malignant smell, the accuracy (and reproducibility in the methodological approach) of dogs' work were, however, insufficient for practical purposes, possibly due to dogs' individual condition and competence, trainers' influences on dogs work, or non-optimal carrier materials.

The use of breath carrier materials providing reversible binding of VOC, such as earloop masks and fleece-filled tubes, and the careful handling of samples all contributed to a high hit rate of dogs. Also, a careful selection of appropriate sampling material is an important factor for successful testing.

The results obtained with an eNose showed a clear differentiation between LC, COPD and healthy breath samples. Despite the low number of analysed samples, an unexpected and significant finding was that carrier materials, specifically fleece earloop masks and fleeces in glass tubes, qualified very well and yielded results via eNose that were comparable to those of directly collected breath. The possibility to collect breath samples with appropriate sampling material, to store them for a limited

period of time, and then to send them to a center where they can be assessed with an eNose provides scope for further studies.

Compared to dog-based screening, the application of an eNose as a screening tool provides the advantage of shorter calibration times, higher diagnostic accuracy, higher reproducibility and better standardization (including common biostatistic analysis).

8. References

1. Biehl, Wiebke; Hattesohl, Akira; Jörres, Rudolf A.; Duell, Thomas; Althöhn, Ulrike; Koczulla, Andreas Rembert; Schmetzer, Helga. VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials. Acta oncologica (Stockholm, Sweden) 2019; 1–9.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: A Cancer Journal for Clinicians 2011; 61(2): 69-90.
3. Durham, A. L.; Adcock, I. M. The relationship between COPD and lung cancer. Lung cancer (Amsterdam, Netherlands) 2015; 90(2): 121-127.
4. Mountain C. Revisions in the international system for staging lung cancer. Chest 1997; 111: 1710-1717.
5. Shim SS, Lee KS, Kim BT, Chung MJ, Lee EJ, Han JH, Choi JY, Kwon OJ, Shim YM, Kim S. Non-small cell lung cancer: prospective comparison of intefrated FDGPET/CT and CT alone for preoperative staging. Radiology 2005; 236(3): 1011-1019.
6. Bach PB, Niewoehner DE, Black WC. Screening for lung cancer: the guidelines for the early detection of cancer 2003; Chest; 123: 83S-88S.
7. Pio, R; Garcia, J; Corrales, L; Ajona, D; Fleischhacker, M; Pajares, M J; Cardenal, F; Seijo, L; Zulueta, J J; Nadal, E; Witt, C; Lozano, M D; Schmidt, B; Montuenga, L M. Complement factor H is elevated in bronchoalveolar lavage fluid and sputum from patients with lung cancer. CA: A Cancer Journal for Clinicians 2010; 19(10): 2665–2672.
8. Edell E, Lam S, Pass H, Miller YE, Stutedja T, Kennedy T, Loewen G, Keith RL. Detection and localization of intraepithelial neoplasia and invasive carcinoma using fluorescence-reflectance bronchoscopy: an international, multicenter clinical trial. J Thorac Oncol 2009; 4: 49-54.
9. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer. CA: A Cancer Journal for Clinicians 2003; 53(1): 27-43.
10. van der Heijden, E H F M; Hoefsloot, W; van Hees, H W H; Schuurbiers, O C J. High definition bronchoscopy: a randomized exploratory study of diagnostic

value compared to standard white light bronchoscopy and autofluorescence bronchoscopy. Respiratory research 2015; 16: 33.

11. Moyer, V A. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Annals of internal medicine 2014; 160(5): 330–338.
12. Horvath I, Lazar Z, Gyulai N, Kollai M, Losonczy G. Exhaled biomarkers in lung cancer. The European Respiratory Journal 2009; 34(1): 261-275.
13. Czitrovszky A, Szymanski W, Nagy A et al. A new method for the simultaneous measurement of particle size, complex refractive index and particle density. Measurement Science and Technology 2002;13:303-308.
14. Machado RF, Laskowski D, Deffenderfer O, Burch T, Zheng S, Mazzone PJ et al. Detection of lung cancer by sensor array analysis of exhaled breath. American Journal of Respiratory and Critical Care Medicine 2005; 171:1286-91.
15. Dragonieri, S; Annema, J T.; Schot, R; van der Schee, M P C; Spanevello, A; Carratú, P; Resta, O; Rabe, K F; Sterk, P J. An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. Lung cancer (Amsterdam, Netherlands) 2009; 64(2): 166-170.
16. Tan, J; Yong, Z; Liam, C. Using a chemiresistor-based alkane sensor to distinguish exhald breaths of lung cancer patients from subjects with no lung cancer. Journal Of Thoracic Disease 2016; 8(10): 2772-2783.
17. Gasparri, R; Santonico, M; Valentini, C; Sedda, G; Borri, A; Petrella, F; Maisonneuve, P; Pennazza, G; D'Amico, A; Di Natale, C; Paolesse, R; Spaggiari, L. Volatile signature for the early diagnosis of lung cancer. Journal Of Breath Research 2016; 10(1): 016007
18. Hubers, A J; Brinkman, P; Boksem, R J.; Rhodius, R J.; Witte, B I.; Zwinderman, A H. Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis. Journal Of Cinical Pathology 2014; 67(8): 707-711
19. Montuschi, P; Mores, N; Trové, A; Mondino, C; Barnes, P J. The electronic nose in respiratory medicine. Respiration, International Review Of Thoracic Diseases 2013; 85(1): 72-84

20. Horvath, G; Järverud, G A K; Järverud, S; Horváth, I. Human ovarian carcinomas detected by specific odor. Integrative Cancer Therapies 2008; 7(2): 76-80.
21. Sonoda H, Kohnoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, Morita M, Watanabe A, Morita M, Kakeji Y, Inoue F, Maehara Y. Colorectal cancer screening with odour material by canine scent detection. Gut 2011; 60: 814-819.
22. Cornu JN, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. European Urology 2011; 59(2): 197-201.
23. McCulloch M, Jezierski T, Broffman M; Hubbard, A; Turner, K; Janecki, T. Diagnostic accuracy of canine scent detection in early - and late - stage lung and breast cancers. Integrative Cancer Therapies 2006; 5(1): 30-39.
24. Ehmann, R; Boedeker, E; Friedrich, U; Sagert, J; Dippon, J; Friedel, G; Walles, T. Canine scent detection in the diagnosis of lung cancer: revisting a puzzling phenomenon. European Respiratory Journal 2012; 39(3): 669-676.
25. Amundsen, T; Sundstrøm, S; Buvik, T; Gederaas, O A; Haaverstad, R. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. CA: A Cancer Journal For Clinicians 2014; 53(3): 307–315.
26. Hackner, K; Errhalt, P; Mueller, M R; Speiser, M; Marzluf, B A.; Schulheim, A; Schenk, P; Bilek, J; Doll, T. Canine scent detection for the diagnosis of lung cancer in a screening-like situation. Journal Of Breath Research 2016; 10(4): 046003.
27. Greulich, T; Hattesohl, A; Grabisch, A; Koepke, J; Schmid, S; Noeske, S; Nell, C; Wencker, M; Jörres, R A; Vogelmeier, C F; Köhler, U; Koczulla, A R. Detection of obstructive sleep apnoea by an electronic nose. The European Respiratory Journal 2013; 42(1): 145–155
28. Koczulla, A. R.; Hattesohl, A.; Biller, H.; Hofbauer, J.; Hohlfeld, J.; Oeser, C.; Gessner, C.; Vogelmeier, C.; Baumbach, J. I.; Wirtz, H.; Jörres, R. A. Krankheiten erriechen? Eine kurze Übersicht über elektronische Nasen. Pneumologie 2011; 7: 401-405.

29. Helton, WS. Canine ergonomics, The science of working dogs. CRC Press/Taylor & Francis 2009
30. Bouton, ME. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. Biological Psychiatry 2002; 52(10): 976–986
31. Robbins, Patrick J.; Ramos, Meghan T.; Zanghi, Brian M.; Otto, Cynthia M. Environmental and Physiological Factors Associated With Stamina in Dogs Exercising in High Ambient Temperatures. Frontiers in veterinary science 2017; 4: 144

9. List of tables

(Parts of) Tables 1, 2, 4, 6, 7, 9, 10, 11, 13 are already published in Biehl et al., Acta oncologica 2019 [1].

10. List of figures

(Parts of) Figures 1, 9, 10, 11, 12, 13 are already published in Biehl et al., Acta oncologica 2019 [1].

Figure 8 is already published in Smith detection The Cyranose® 320 ENose User's Manual part number 11-60001, page 10.

11. Abbrevations

LC	Lung cancer
COPD	Chronic obstructive pneumological disease
CT	Computer tomography
LDCT	Low-dose computer tomography
VOC	Volatile organic compounds
eNose	Electronic nose
SCLC	Small cell lung cancer
NSCLC	Non small cell lung cancer
LDA	Linear discriminant analysis
BMI	Body Mass Index
CVV	Cross validation value
PPV	Positive predictive value
NPV	Negative predictive value

12. List of publications

12.1 Original studies

Wiebke Biehl, Akira Hattesohl, Rudolf A. Jörres, Thomas Duell, Ulrike Althöhn, Andreas Rembert Koczulla & Helga Schmetzer: **VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials.** Acta Oncologica. 2019 July; 1–9

12.2 Congress contributions

Abstract submitted: Biehl, Wiebke; Hattesohl, Akira; Jörres, Rudolf A.; Duell, Thomas; Althöhn, Ulrike; Koczulla, Andreas Rembert; Schmetzer, Helga. **VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials.** 7th Immunotherapy of Cancer Conference (ITOC7) April 2-4 2020, Munich, Germany.

Abstract submitted: Biehl, Wiebke; Hattesohl, Akira; Jörres, Rudolf A.; Duell, Thomas; Althöhn, Ulrike; Koczulla, Andreas Rembert; Schmetzer, Helga. **VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials.** 46th Edition of the EBMT Annual Meeting, March 22-25 2020, Madrid, Spain.

12.3 Online and printed contributions

Hoffnung für Patienten mit Lungenkrebs, Asthma und COPD – Wie eine elektronische Nase Krankheiten aufspürt. Focus, 27.06.2014

Atemluft-Test soll Magenkarzinome „riechen“ – und könnte als Screening taugen. Medscape, 13.05.2015

Atemluft-Test soll Magenkarzinome „riechen“ – und könnte als Screening taugen. Medscape, 09.10.2017

Hier riecht’s nach Tumor. DocCheck-News, 21.03.2018

12.4 TV contributions

24.10.2017: Contribution in NDR Visite

27.04.2015: Contribution in Wissensmagazin X:enius (Arte)

Under progress: Contribution for a documentary for Swiss television and 3sat

14. Danksagung

An erster Stelle möchte ich mich bei Frau Prof. Dr. Helga Schmetzer für die langjährige umfangreiche und intensive Betreuung meiner Dissertation bedanken sowie für das angenehme und freundschaftliche Verhältnis, in dem wir arbeiteten. Sie war immer als Ansprechpartnerin bei Fragen und Anregungen für mich da und hat in der phasenweise sehr schwierigen Studie stets zum Weitermachen motiviert. Außerdem möchte ich mich herzlich für die Mitarbeit an unserer Veröffentlichung bedanken. Frau Prof. Dr. Helga Schmetzer betreute mich mit viel Herz und konnte mir von Beginn der Studie bis zum heutigen Tag an immer die Sicherheit vermitteln, das Projekt irgendwann erfolgreich zu beenden.

Prof. Dr. med. Rembert Koczulla aus der Universitätsklinik Marburg danke ich für die Teilnahme an unserer Studie und die tolle und unkomplizierte Zusammenarbeit. Aus seinem Team danke ich ganz besonders MTA Frau Ursula Boas für die freundliche Kommunikation und Mitarbeit an unserer Studie sowie dem Bioinformatiker Herrn Akira Hattesohl für die große Unterstützung in Bezug auf die Arbeit mit der elektronischen Nase, die statistische Auswertung und sein stets offenes Ohr bei Fragen.

Bei PD Dr. R. Jörres bedanke ich mich für die jahrelange Mitwirkung am Projekt, die Hilfe bei der statistischen Auswertung, die konstruktive Kritik und die Vermittlung von neuen Kontakten.

Dr. med. Thomas Duell aus der Lungenfachklinik in Gauting danke ich für die Einarbeitung und interdisziplinäre Vermittlung im Krankenhaus, die Bereitstellung eines Arbeitsplatzes und von Patientenproben sowie die Mitwirkung am Projekt.

Bei allen Hundeführern und -trainern, ganz besonders aber bei dem komplett ehrenamtlich arbeitenden Team aus Gießen, möchte ich mich hiermit herzlich bedanken für ihre Teilnahme, für die Bereitstellung ihrer Zeit und ihrer Hunde.

Ich bedanke mich bei allen Mitarbeitern der mitwirkenden Abteilungen der Asklepios Klinik München Gauting sowie der Universitätsklinik Marburg für die Hilfe vor Ort und die Überlassung von klinischen Daten. Ein ganz herzliches Dankeschön gilt natürlich

auch allen Patienten und gesunden Probanden, die sowohl für das Training der Hunde als auch für die doppelblinde Studie bereit waren, Atemproben zu spenden.

Der Stiftungsleiterin der Innovationsstiftung Ulrike Sauer danke ich für die Initiierung des ersten Teils des Projekts und die Bereitstellung eines Appartements während meiner Tätigkeit in München.

Ein besonderer Dank geht an meinen Ehemann für die Motivation und Unterstützung während der Jahre meiner Promotion, aber auch meinen Eltern für die Ermöglichung dieses Schrittes, sowohl finanzieller als auch psychologischer Natur.

Ich widme diese Arbeit meinem verstorbenen Großvater Prof. Dr. Boi Böle Bernhard Biehl, der mich von Beginn der Arbeit mit seinen interessierten, aber auch kritischen Kommentaren voran brachte und leider das endgültige Resultat nicht mehr lesen konnte, sowie meinem ebenfalls bereits verstorbenen ersten Hund Herr Nilsson, ohne den ich vermutlich niemals über dieses Thema promoviert hätte.

15. Eidesstaatliche Versicherung

Ich, Wiebke Ingalisa Biehl, erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

VOC Pattern Recognition of Lung Cancer: a Comparative Evaluation of Canine and eNose- Based Strategies Using Different Sampling Materials

selbstständig verfasst, mich außer den unten angegebenen keiner weiteren Hilfen bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annährend übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich sammelte Atemproben für den methodischen Teil der Studie, initiierte das Hundeteam für den zweiten Teil der Studie, war verantwortlich für die Datenerfassung, Auswertung und statistische Aufarbeitung, war an der Entwicklung des Studiendesigns beteiligt, erstellte die gleichnamige Publikation (Biehl et al., Acta oncologica 2019 [1]) mit Hilfe der unten genannten Autorenbeiträge und wurde von Helga Schmetzer betreut.

PD Dr. Rudolf Jörres unterstützte die Studie, trug Teile der statistische Auswertungen und viele Denkanstöße für die Diskussion bei.

PD Dr. Thomas Duell war an der Entwicklung des Studiendesigns beteiligt und unterstützte das Projekt mit Bereitstellung von Patienten für die Abnahme von Exhalat und Erfassung ihrer klinischen Daten.

Akira Hattesohl arbeitete mit an der statistischen Datenanalyse der eNose-Ergebnisse und an der eNose-spezifischen inhaltlichen Korrektur der Publikation.

Ulrike Althöhn war im Hauptteil der Studie für die Hundeausbildung verantwortlich.

Prof. Dr. Andreas Rembert Koczulla war an der Entwicklung des Studiendesigns beteiligt, war insbesondere für die eNose-bezogenen Teile der Studie verantwortlich,

unterstützte die Studie mit Atemproben, klinischen Daten der Patienten, diskutierte und überprüfte die Publikation.

Prof. Dr. Helga Schmetzer schrieb das Ethikvotum, unterstützte, leitete und überwachte das gesamte Projekt und die Arbeit von Wiebke Biehl, sowohl die Publikation als auch die Dissertation betreffend.

Ich erkläre des Weiteren, dass die hier vorliegende Dissertation nicht in gleicher oder ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Harsefeld, den 26.02.2021	Wiebke Biehl
Ort, Datum	Unterschrift des Doktoranden