Journal of Breath Research

PAPER

A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD

A Gaida[1,2,8], O Holz[3,8,9], C Nell[4], S Schuchardt[5], B Lavae-Mokhtari[1], L Kruse[1], U Boas[6], J Langejuergen[7], M Allers[7], S Zimmermann[7], C Vogelmeier[4], A R Koczulla[4] and J M Hohlfeld[3]

1 Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Clinical Airway Research, Hannover, Germany
2 Leibniz Universität Hannover, Institute of Inorganic Chemistry, Research Group Analytical Chemistry, Hannover, Germany
3 Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Clinical Airway Research, Member of the German Center for Lung Research (BREATH), Hannover, Germany
4 Philipps-Universität Marburg, Pneumology, Member of the German Center for Lung Research (UGMLC), Marburg, Germany
5 Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Bio- and Environmental Analytics, Hannover, Germany
6 Philipps-Universität Marburg, Pneumology, Marburg, Germany
7 Leibniz Universität Hannover, Institute of Electrical Engineering and Measurement Technology, Hannover, Germany
8 Contributed equally.
9 Author to whom any correspondence should be addressed. Department of Clinical Airway Research, Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Clinical Airway Research, Member of the German Center for Lung Research (BREATH), Hannover, Germany.

E-mail: olaf.holz@item.fraunhofer.de

Abstract

There is increasing evidence that breath volatile organic compounds (VOC) have the potential to support the diagnosis and management of inflammatory diseases such as COPD. In this study we used a novel breath sampling device to search for COPD related VOCs. We included a large number of healthy controls and patients with mild to moderate COPD, recruited subjects at two different sites and carefully controlled for smoking.

222 subjects were recruited in Hannover and Marburg, and inhaled cleaned room air before exhaling into a stainless steel reservoir under exhalation flow control. Breath samples (2.5 l) were continuously drawn onto two Tenax® TA adsorption tubes and analyzed in Hannover using thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). Data of 134 identified VOCs from 190 subjects (52 healthy non-smokers, 52 COPD ex-smokers, 49 healthy smokers, 37 smokers with COPD) were included into the analysis.

Active smokers could be clearly discriminated by higher values for combustion products and smoking related VOCs correlated with exhaled carbon monoxide (CO), indicating the validity of our data. Subjects from the study sites could be discriminated even after exclusion of cleaning related VOCs. Linear discriminant analysis correctly classified 89.4% of COPD patients in the non/ex-smoking group (cross validation (CV): 85.6%), and 82.6% of COPD patients in the actively smoking group (CV: 77.9%). We extensively characterized 134 breath VOCs and provide evidence for 14 COPD related VOCs of which 10 have not been reported before.

Our results show that, for the utilization of breath VOCs for diagnosis and disease management of COPD, not only the known effects of smoking but also site specific differences need to be considered. We detected novel COPD related breath VOCs that now need to be tested in longitudinal studies for reproducibility, response to treatment and changes in disease severity.

Introduction

Our understanding of breath volatile organic compounds (VOC) has greatly advanced in the past years [1–4]. There is increasing evidence that specific breath VOC patterns not only exist for different lung diseases, but also for systemic diseases and cancer. Specific pathogen related VOC signals promise earlier

and more specific diagnosis and treatment of infections [5–7]. To assess lung and systemic inflammatory processes in chronic obstructive pulmonary disease (COPD), different devices, methodologies and study designs were used in the past [2, 7–20]. In these studies it was possible to discriminate the breath of COPD patients from healthy subjects, from patients with asthma or tumor patients. However, despite these intensive efforts breath research in COPD is still in the exploratory phase of gathering experience with different sampling and analysis strategies to gather data of VOCs and VOC patterns with potential clinical value. An ERS taskforce and one by the international association of breath research (IABR) [21] were set up in 2013 to work on standardization.

For one electronic nose (eNose) device there is data available for comparison between sites [22] and in one eNose study independent patient groups from a different site were used for cross validation (CV) [23]. We are aware of only two studies that used independent validation cohorts for samples that were analyzed by GC-MS [24, 25].

In this study we collected breath samples of COPD patients and controls in two German centers for lung research (Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH) and Universities Giessen and Marburg Lung Centre (UGMLC)). This way we were able to perform one of the largest exhaled breath COPD studies so far, including 190 patients and controls. It also allowed to specifically test the influence of study site. As active smoking is known to have a profound effect on the exhaled breath VOC composition [26] we carefully controlled for the smoking bias by including smokers with COPD, ex-smokers with COPD, healthy smokers and healthy non-smokers into the study. Breath samples were loaded directly onto Tenax® TA tubes during tidal breathing under flow control without the intermediate storage in bags. All samples were centrally analyzed by TD-GC-MS at the analytical laboratory of the Hannover study site.

Methods

Subjects
At the Hannover site 31 smokers with moderate COPD, 30 ex-smokers with COPD, 29 healthy smokers and 29 non-smokers were enrolled into this study. At the Marburg site the respective numbers of subjects were 13, 26, 33, and 31. Due to technical reasons (e.g. insufficient loading of Tenax® tubes, undetected leakage in the TD–GC transfer line system, or other instrument failures during the analysis), a total of 12 (Hannover) and 13 (Marburg) breath samples could not be evaluated. Patients were recruited in random order and not by group to avoid a potential analysis bias. Table 1 lists the demographics of all subjects that were included into the final analysis. Subjects had to be free of exacerbations or acute infections within four weeks prior to the study day. The study was conducted

in accordance with Good Clinical Practice and the Declaration of Helsinki. Subjects gave their written informed consent. The study was approved by the Ethical Committees of Hannover Medical School (Nr. 6490) and Marburg University (AZ 179/12).

Study design
The study was conducted at two sites in Germany from December 2013 to August 2015. Subjects visited the site on a single day and were asked to abstain from eating, drinking (except water) and smoking at least 2 h prior to the visit. After providing informed consent, subjects underwent a thorough physical examination. A lung function test was performed and a blood sample was taken to provide a serum sample, which was stored at -80 °C. The COPD Assessment Test to inquire QoL (CAT) score was assessed by the standard questionnaire. Next we analyzed exhaled carbon monoxide (Smokelyzer, Bedfont, Kent, UK) to verify the smoking status and nitric oxide (NO, NIOX Mino, Aerocrine, Bad Homburg, Germany). The analysis of breath VOCs was performed as described in detail below. In Hannover the collection of breath VOCs was performed in a temperature and humidity controlled room. In addition to sampling breath VOCs onto Tenax® TA tubes we used an electronic nose (Cyranose 320, Sensigent, Baldwin Park, USA) and an IMS detector (BioScout, B&S Analytik, Dortmund, Germany) in a subset of patients in Marburg and used a close gas loop high resolution GC-IMS device [27–29] as well as an ultra-high sensitive TD-GC-APCI-MS [30] in a subset of subjects in Hannover. The data of these subgroup analyses and their comparison with the GC-MS data will be reported separately [31].

Collection of breath samples for GC-MS analysis
The volunteer subjects inhaled through an A2 carbon filter (Dräger, Lübeck, Deutschland; adapted from standard operating procedure of P Brinkman, AMC University of Amsterdam [9]) and exhaled into a stainless steel tube sampling reservoir (50 cm length, 4 cm inner diameter) with flow restrictor at the end of the tube. In Hannover an inhalation/exhalation valve from Hans Rudolph (Shawnee, USA), in Marburg from Intersurgical (St. Augustin, Germany) was used. The first 3 min of each measurement were used to remove acutely accumulated environmental VOCs from the lung and to condition the stainless steel reservoir. Following this, the subjects continued exhaling into the reservoir for a five minute breath collection. Subjects exhaled against a small resistor and were asked to maintain an exhalation pressure between 3 or 6 mbar by visual control of a manometer.

Loading of two Tenax® TA tubes (Perkin Elmer, Waltham, USA) from the reservoir was achieved by using an in-house vacuum line (Hannover) or vacuum pump (Marburg) with separate pre-calibrated restrictors for each tube that allowed 500 ml min^{-1} (variation <2% between tubes) to pass through the

Table 1. Subjects demographics.

	Non/Ex-smoker				Smoker				p ANOVA	
	Healthy		COPD		Healthy		COPD			
	Han	Mar	Han	Mar	Han	Mar	Han	Mar	Han	Mar
n	27	25	28	24	24	25	26	11		
Gender (f/m)	16f/11m	14f/11m	12f/16m	6f/18m	16f/8m	12f/13m	13f/13m	2f/9m		
Age (years)	38 ± 13	32 ± 13	63 ± 05***	65 ± 07***	47 ± 13	40 ± 14	61 ± 07***	61 ± 18**	***	***
Height (cm)	173.9 ± 8.4	174.9 ± 9.7	170.3 ± 9.7	175.7 ± 9.5	169.4 ± 8.9	174.6 ± 11.4	172.6 ± 11.3	172.0 ± 10.8		
Weight (kg)	71.1 ± 14.4	72.5 ± 11.6	86.7 ± 17.5***	86.3 ± 15.9*	74.3 ± 17.7	79.4 ± 17.2	79.1 ± 14.9	88.1 ± 23.8	**	*
BMI	23.4 ± 3.6	23.6 ± 2.5	29.8 ± 5.1***	28.0 ± 4.9*	25.7 ± 4.9	25.9 ± 4.8	26.4 ± 3.6	29.6 ± 6.6*	***	**
FVC (L)	4.6 ± 1.0	4.8 ± 1.1	4.0 ± 1.0	3.4 ± 1.1**	4.2 ± 0.7	4.7 ± 1.3	4.2 ± 1.3	3.5 ± 0.9**		***
FVC(%pred.)	110.1 ± 12.5	108.9 ± 13.3	116.0 ± 26.5	85.7 ± 18.6***	116.2 ± 17.0	108.1 ± 12.7	115.5 ± 16.6	89.8 ± 21.5***	***	***
FEV1 (L)	3.6 ± 0.7	4.0 ± 0.9	1.9 ± 0.7***	2.0 ± 0.6***	3.2 ± 0.6	3.8 ± 1.0	2.3 ± 0.9***	1.9 ± 0.6***	***	***
FEV1 (% pred.)	103.2 ± 11.5	105.1 ± 10.5	67.3 ± 23.6***	65.5 ± 17.3***	105.0 ± 11.6	104.3 ± 11.4	79.2 ± 18.0***	58.4 ± 14.2***	***	***
FEV1/FVC (%)	79.5 ± 5.5	82.4 ± 6.7	46.5 ± 11.7***	58.6 ± 10.1***	76.9 ± 5.8	80.9 ± 7.8	55.9 ± 10.9***	53.5 ± 11.9***	***	***
PY	0(0)	na	$49.9(28.9)^{\#\#\#}$	na	31.4 (31.0)	na	$46.0(18.2)^{\#}$	na		
GOLD (n)			9/11/7/1	7/14/1/2			14/10/2/0	1/3/6/1		
CAT score			13.0 (8.5)	18.3 (12.9)	13.0 (8.5)		13.0 (8.5)	24.0 (3.5)		
FeNO (ppb)	17.0 (9.0)	14.3 (11.0)	21.5 (13.9)	18.3 (12.9)	8.8 (5.1)	10.7 (3.7)	9.5 (5.4)	5.7 (4.5)***	***	***
CO	2.0 (2.0)	2.6 (1.2)	2.0 (2.0)	3.9 (1.9)	10.5 (4.5)	8.3 (12.5)	13.5 (13.8)	27.0 (16.6)***	***	***
COHb	0.9 (0.4)	1.0 (0.2)	0.9 (0.4)	1.2 (0.3)	2.3 (0.7)	1.9 (2.0)	2.8 (2.2)	4.9 (2.7)***	***	***

Han = Hannover site, Mar = Marburg site, BMI = body mass index, GOLD = numbers of patients with stages 1/2/3/4, COPD lung function was assessed post bronchodilator, CAT = COPD assessment test, FeNO = level of exhaled nitric oxide, CO = exhaled carbon monoxide, COHb = calculated carboxyhemoglobin from exhaled CO analysis, mean ± SD and median (interquartile range) are given, ANOVA = analysis of variance, all groups (healthy versus COPD post hoc-Newman–Keuls test).

*** $= p < 0.001$, ** $= p < 0.01$, * $= p < 0.05$, ### $= p < 0.001$, ## $= p < 0.01$, # $= p < 0.05$ MWU test healthy versus COPD. ANOVA for NO, CO, COHb for log transformed data.

tube. An additional tube was used to collect 2.5 l of room air close to the study subject using identical sampling conditions. Due to a temporary lack of thermal desorption tubes (in a period of TD-GC-MS instrument maintenance and repair) the room air control could not be sampled in some subjects at both sites.

VOC analysis

Tenax® TA tubes were conditioned in Hannover (250 °C, 5 min) and either stored or shipped to Marburg for sampling. Tubes were tightly closed with Swagelok Caps containing PTFE fittings (polytetrafluorethylen, Swagelok®, Solon, USA) for storage and shipment. The time between sampling and analysis was generally less than 1 week for Marburg samples and less than 3 d for Hannover samples.

Breath VOCs were analyzed by TD-GC-MS and identified in Hannover (for technical specifications see online supplement 1 (stacks.iop.org/JBR/10/026006/mmedia)) using Enhanced Chemstation Software Rev. E.02.02 (Agilent Technologies, Santa Clara, USA). Compound identification was performed by retention time, reference compounds and by library search using the National Institute of Standards and Technology (NIST) Mass Spectral Search Program Version 2.0 (NIST, Boulder, USA). The detailed parameters for the analysis are listed in table S1 (online supplement) and examples for chromatograms from the different subject groups as well as for room air are also provided in the online supplement (figures S1 and S2). 134 VOC were quantified using MS response of specific target ions which most commonly matched with m/z signals of highest intensity in the respective mass spectrum. To avoid faulty discrimination between signals of same m/z ratio in the chromatogram (i.e. m/z 43 in acetic acid and ethyl acetate), other more specific target ions were selected for quantification.

Identification of VOCs using reference compounds

A subset of 30 compounds was chosen for external standard validation and calibration (table 2). The reference compounds chosen for external standard measurements included linear alkanes, alcohols, aldehydes, aromatic compounds and terpenes all of which were detected in more than 90 percent of the subjects' samples. Pure compounds were solved in methanol yielding concentrations between 1 and 100 ng μl^{-1} and manually adsorbed onto Tenax® TA tubes. Therefore 1 μl of external standard solution was injected onto the top of the Tenax® TA surface using micro syringes for GC injection. To avoid loss of insufficiently retained VOCs prior to analysis during TD dry purge and flushing procedures, the tubes were immediately brought into a stream of dry nitrogen carrier gas (200 ml min^{-1}, 2 min) to completely adsorb the VOC molecules to the Tenax® TA. Samples were injected using the same TD-GC-MS parameters as breath samples. Retention time and mass spectrum were used to validate NIST library search suggestions.

Except for glutaraldehyde (No. 12 second best NIST match vinyl acetate in table 2), in all tested 30 cases retention times matched and NIST identification was confirmed. Furthermore reference compounds were used to assess concentration ranges of breath compounds and their reproducibility with respect to quantification (standard deviation/mean for five-fold experiment <5%). Single ion counts between 0.1 and 10 ng represented typical response values for compounds with normal abundance in 2.5 l of breath.

Data processing

The raw data of two Tenax® tubes were taken for the analysis without normalization or correction with room air sample data. We found that the sum of total ion current (TIC) was not suited for normalization as the chromatograms included drastically varying peaks e.g. related to cleaning of valves or to Tenax® degradation. The raw data of two simultaneously loaded Tenax® tubes was first compared to identify outliers. The mean value of the single ion response of the two tubes was calculated and log transformed as this data was not normally distributed. Subjects showing hardly detectable VOC peaks due to improper sampling, the typical Tenax® matrix decomposition VOC pattern, and most known indoor contaminations (e.g. cleaning reagents) were at this point excluded. Thus 101 VOCs of 190 subjects remained in the database for the final statistical data analysis.

Statistics

Log transformed data of breath VOCs are displayed as mean and standard error (SE) of the mean in the respective figures. Univariate comparison was performed by t-test. Demographic data were analyzed by ANOVA and post-hoc analysis was performed by the Newman–Keuls test. For correlations the Pearson correlation coefficient was given. Linear discriminant analysis (LDA) was used to determine the major VOCs that discriminate between patient groups (SPSS 20, IBM, Armonk, USA). The models were computed using stepwise inclusion. Cross-validation was performed by the leave-one-out method.

Results

Demographics, comparison between sites

The overall COPD severity as indicated by the CAT score was milder in Hannover, but there were more GOLD 3 patients among the ex-smokers. Among the actively smoking COPD patients the highest mean CAT score and the highest CO levels were found in the COPD patients recruited in Marburg. There were also more GOLD 3 and 4 patients in this group (table 1). COPD patients were significantly older than healthy control subjects and had a larger BMI. Among the COPD patients recruited in Hannover were more women. Active smoking in COPD patients and healthy smokers resulted in significantly higher values of exhaled CO

Table 2. List of breath VOCs.

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related (univariate) all[22]	COPD related (univariate) H[23]	COPD related (univariate) M[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/smoking status[28]
1	Acetaldehyd	1.6	100	93	92	100	100	89	93	91	81															x			
2	Ethanol	2.1	63	79	67	74	72	56	53	55	79					x						1							
3	Isoprene	2.4	96	100	96	100	100	96	100	100	94							M								x			
4	Acetone	2.5	100	100	100	100	100	96	100	100	100							M				>100				x			
5	Acetonitrile	2.7	59	75	96	100	76	81	87	100	6				x											x			
6	Dimethyl sulfide	2.8	100	100	100	100	100	89	100	100	0															x			
82	1,3-Dichlor-2-propanol	2.9	19	18	8	7	24	41	30	73	37					x			M										
7	Acetic acid methyl ester	3.0	96	86	71	63	92	81	83	73	39				x											x			
83	Carbon disulfide	3.0	0	11	4	0	0	11	10	18	13		x					M											
8	1,3-Cyclopentadiene	3.1	0	0	50	41	0	0	30	27	0		x	x								1–5							
9	1-Propanol	3.5	59	61	63	59	68	48	50	82	83					x										x			
84	1-Hexene	3.5	0	7	63	63	4	4	37	91	15				x											x	x		
10	n-Hexane	3.6	7	14	54	48	20	11	37	64	68				x	x		M								x			
85	3/4-Methyl-2-pentene	3.9	0	4	79	74	0	7	43	100	1				x											x			
11	Dimethyl selenide	3.9	26	25	29	4	16	19	13	9	0		x																
12	**Vinyl acetate**	4.2	78	96	96	96	88	96	80	100	57							M					N	N	N/S	x	x		x
13	2-Methyl furane	4.3	0	7	88	78	8	11	57	100	2				x											x			
14	**Butanone**	4.5	22	82	71	78	60	85	63	100	57					x						0.25–2	N	N	S	x	x		
15	Ethyl acetate	4.7	4	7	8	7	16	4	13	18	69		x			x		H				n.a				x			
87	2,4-Hexadiene	4.8	0	0	63	63	0	0	43	100	0			x	x								S		S	x			
88	4-Methyl-1,3-pentadiene	4.9	0	0	71	74	0	0	47	100	1			x	x								S	S	S	x			
16	1,3-Cyclohexadiene	5.1	11	25	96	89	36	15	77	100	1				x								S		S	x			
17	1,4-Cyclohexadiene	5.1	26	18	92	89	36	19	80	100	1				x								S		S	x			
89	3-Methyl hexane	5.4	7	11	4	15	16	4	7	27	9		x																

(Continued)

Table 2. (*Continued*)

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related (univariate) all[22]	COPD related (univariate H)[23]	COPD related (univariate M)[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/smoking status[28]
19	**Benzene**	5.8	96	96	100	96	96	96	100	100	100				x	x		H				0.2–2	S		S	x			x
20	Heptane	6.0	67	68	92	93	76	67	87	100	38				x	x					x	n.a		N					
18	**Acetic Acid**	6.1	96	100	100	96	100	96	100	100	43													N		x	x		
90	2-Methyl-1/2-hexene	6.2	0	0	0	0	0	0	0	0	0		x																
91	2-Ethyl furane	6.6	0	4	83	81	0	0	53	100	1				x								S		S	x			
21	Allyl methyl sulfide	6.7	63	75	79	67	52	44	63	91	1														S	x			
22	2-Pentanone	6.9	100	100	100	100	100	89	97	100	4				x							1			S	x			
23	2,5-Dimethyl furane	6.9	4	0	96	89	4	11	80	100	0				x										S	x			
24	Pentanal	7.0	52	46	46	41	48	7	33	27	8													N		x			
92	2,4-Dimethyl furane	7.0	4	0	96	89	4	11	80	100	0				x											x			
25	Methyl cyclohexane	7.0	11	18	25	30	24	11	27	36	19		x													x			
26	1-Methyl thiopropane	7.2	74	82	67	63	44	63	63	73	0												S		S	x			
93	Methyl methacrylate	7.3	4	4	4	22	8	7	3	9	2		x													x			
27	2-Vinyl furane	7.5	0	0	46	59	0	0	27	91	0				x								S		S	x			
28	(E)-1-Methyl thiopropene	7.5	52	61	29	67	52	37	40	73	0			x												x			
94	Unidentified C7-dialkene	7.7	4	4	79	81	4	4	47	100	0				x								S		S	x			
95	Unidentified C7-dialkene	7.8	4	7	83	81	8	7	50	100	1				x								S		S	x			
30	(Z)-1-Methyl thiopropene	7.9	93	96	88	81	64	67	67	91	0															x			
96	Propanoic acid	7.9	15	32	33	44	16	22	23	36	0															x			
29	1-Methyl-1,4-cyclohexadiene	8.1	7	7	96	81	8	7	67	100	1				x								S		S	x			
31	1,3-Cycloheptadiene	8.3	4	11	71	81	12	0	37	100	0				x								S		S	x			
97	1,3-Dimethyl cyclohexane	8.5	7	11	17	22	8	7	13	27	3		x																
98	Unidentified C7-dialkene	8.6	7	18	75	78	16	4	37	91	0				x								S		S	x			
99	3/4-chloro-Octane	8.8	22	25	71	67	20	11	30	82	10				x											x			
32	**Toluene**	8.8	100	100	100	100	100	96	100	100	99				x	x						1–5	N/S	N	S	x			x

(*Continued*)

Table 2. (*Continued*)

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related all (univariate)[22]	COPD related H (univariate)[23]	COPD related M (univariate)[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/smoking status[28]
33	Octane	8.9	93	93	100	96	80	93	93	100	91				x	x						0.1–0.5				x			
100	1,4-Dimethyl cyclohexane	9.3	0	0	0	0	0	0	0	0	0		x																
101	2-Ethyl-5-methyl furane	9.6	0	0	83	81	0	0	50	100	1			x	x								S		S	x			
34	Hexanal	9.9	93	82	92	81	60	59	63	91	69					x						0.5–2.5				x			
102	1,3,5-Trimethyl 1H-Pyrazole	10.0	0	0	29	44	0	0	23	64	0		x	x	x											x			
35	Tetrachlorethylene	10.0	33	18	13	11	36	7	17	0	1		x																
103	2,3-Dimethyl-1,3-cyclohexadiene	10.7	0	0	58	59	0	0	27	91	0			x	x								S		S	x			
36	1,2-Dimethyl-1,4-cyclohexadiene	10.7	0	0	58	59	0	0	27	82	0			x	x								S		S	x			
37	1-Nonene	11.7	63	61	79	74	44	56	67	100	61				x	x							S		S	x			
38	Ethyl benzene	11.7	85	93	96	96	80	74	100	100	99				x	x							S		S	x			
39	Nonane	11.8	100	100	100	100	88	96	93	100	97	x				x						>0.5	S		S	x			
40	**m,p-Xylene**	11.9	85	93	96	96	80	74	100	100	99				x	x						0.1–1	S		S	x			x
104	**1,6-Dimethyl-1,3,5-heptatriene**	12.0	0	7	71	74	0	19	40	100	26			x	x								N/S	N	N/S	x			x
41	o-Xylene	12.7	26	50	96	93	36	33	57	100	87				x	x						0.1	S		S	x		x	
42	Styrene	12.7	26	50	96	89	60	59	77	100	59				x	x							S		S	x			
43	Heptanal	12.8	93	86	96	89	64	63	67	100	65						M									x			
44	α-Pinene	13.3	100	100	100	100	100	89	97	100	69											0.1–1				x			
45	Camphene	13.9	22	25	54	67	20	4	33	91	1				x								S		S	x			
46	1-Decene	14.3	44	39	79	78	40	52	77	100	16				x											x			
105	2,2,4,6-Pentamethylheptane	14.3	81	75	92	70	76	70	77	91	74					x				x						x			
47	Decane	14.5	81	75	92	70	76	78	80	100	77								M	x		0.1				x			
48	**1-Ethyl-3-methyl benzene**	14.5	26	61	92	89	28	33	60	100	87				x	x							S		S	x			x
49	Hexanoic acid	14.6	15	14	17	19	56	59	47	82	33					x		M	M				N/S	N	N/S	x			
50	β-Pinene	14.8	100	100	100	100	100	89	93	100	35											0.1–1				x			
106	2,7-Dimethyl-2,6-octadiene	14.8	96	100	96	96	72	74	77	100	7															x	x		

(*Continued*)

Table 2. (Continued)

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related (univariate) all[22]	COPD related (univariate H)[23]	COPD related (univariate M)[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/ smoking status[28]
107	2,6-Dimethyl-2-6-octadiene	14.8	96	100	96	96	68	74	77	100	18															x			
51	2-Pentyl furane	15.0	41	29	46	33	56	26	33	73	13				x								N		N	x	x		
52	**Phenole**	15.2	100	100	100	100	100	96	100	100	100						A					>1	N		N	x	x		
53	Benzaldehyde	15.3	100	100	100	100	100	96	100	100	100	x				x	**A**									x			
109	Trimethylbenzene/Mesitylene	15.3	100	100	100	100	100	96	100	100	100				x		M									x			
108	2,2,4,4-Tetramethyloctane	15.4	78	96	88	81	96	89	80	100	97															x			
55	3-Carene	15.6	100	100	96	100	96	78	67	100	47															x			
54	Octanal	15.6	100	100	96	100	96	89	80	100	93	x					M					0.1–1				x			
110	1-p-Menthene	15.6	85	86	96	93	72	48	73	100	15				x											x			
56	2-Ethyl hexanol	16.1	100	100	100	100	96	96	93	100	97					x	M	M		x		0.1				x			
57	D-Limonene	16.1	100	100	100	100	100	96	100	100	43				x							0.5–2.5			S	x			
58	o/p-Cymen	16.2	96	100	100	96	92	89	83	100	39												S		S	x			
59	1,4-Dichloro benzene	16.4	7	0	0	4	0	0	3	9	0		x																
60	Eucalyptol	16.4	89	93	96	81	72	56	60	82	27															x			
61	1-chloro-Heptane	16.8	48	43	58	56	24	11	17	45	23								H			n.a.				x			
62	γ-Terpinene	16.9	63	64	38	44	24	15	17	55	1											<0.1				x			
63	Undecane	17.1	96	100	100	96	76	78	70	100	89							M	H			0.1–0.5				x			
111	Dihydromyrcenol	17.2	89	79	88	81	32	19	20	27	43								H							x			
65	Linalyl butyrate	17.6	89	75	92	78	24	11	7	9	13															x		x	
112	2-Hydroxy benzaldehyde	17.6	70	82	83	78	92	74	77	91	59						A						S	S		x			
113	**m/p-Cresol**	17.7	44	50	71	48	48	7	30	36	5																	x	x
64	3,7-Dimethyl-3-octanol	17.7	100	100	100	100	100	96	100	100	99															x			
114	2-Methyl-1-propenyl benzene	18.0	19	11	8	19	12	7	0	18	4		x															x	
115	Tetrahydro-2H-Pyran-2-one	18.1	85	82	88	85	76	89	67	100	72					x	A												
66	Nonanal	18.2	100	100	100	100	100	96	97	100	100	x				x	**A**									x		**x**	

(Continued)

Table 2. (*Continued*)

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related (univariate all)[22]	COPD related (univariate H)[23]	COPD related (univariate M)[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/smoking status[28]
67	Acetophenone	18.3	100	100	100	100	100	96	100	100	100	x				x	A		H										
68	4-Methyl benzaldehyde	18.5	89	75	79	63	4	0	0	0	10						H	H	H							x			
69	2-Methyl benzofurane	19.0	44	46	13	22	24	0	7	0	13				x		H	H								x			
117	1-chloro-Nonane	19.4	56	64	63	52	32	30	27	45	50					x										x			
116	Octanoic acid	19.4	22	18	17	33	48	56	43	82	43					x		M	M						S	x			
70	Dodecane	19.5	100	100	100	100	76	74	70	100	81											0.1–1				x			
118	1-Nonanol	19.7	67	50	58	56	40	30	37	73	29															x			
120	3,7-Dimethyl-1-octanol	19.7	63	79	71	78	44	19	23	64	39															x			
119	1,7-Diaminoheptane	19.8	22	25	33	52	0	7	3	18	0								H							x			
71	Benzoic Acid	19.9	63	75	42	63	80	81	63	91	82	x				x		M								x			
72	Menthol	20.1	78	86	83	85	44	52	57	82	47				x			M								x			
121	b-Linalool	20.2	19	25	42	44	0	11	3	27	2				x				H							x			
122	Acetic Acid phenylmethyl ester	20.2	22	46	46	48	20	19	17	45	11															x			
73	Decanal	20.6	100	100	100	100	100	96	100	100	100	x				x	A					0.5–5							
123	Unidentified C12-Ketone	20.6	78	89	92	85	80	74	70	100	81					x													
125	2-Propyl-1-heptanol	20.6	100	100	100	100	100	96	97	100	100					x	M									x			
124	a-Terpineol	20.6	22	36	25	30	4	0	0	18	1								H										
74	Citronellol	21.0	67	64	58	67	20	7	13	0	9								H							x			
126	**Linalyl acetate**	21.5	89	96	88	89	24	11	3	9	9								**H**							x		x	
127	**Tridecane**	21.8	100	100	100	100	88	85	77	100	74											0.1–0.5				x	x		x
128	1-Decanol	22.0	59	71	75	74	60	33	47	82	20															x			
75	2-Phenoxy propanone	22.0	96	96	100	100	72	74	60	82	83											0.1–5				x			
76	2-Phenoxy ethanole	22.0	100	100	100	100	100	93	93	100	98					x	A												
77	1-Phenoxy-2-propanol	22.3	100	96	100	100	96	74	73	73	65					x	A		**H**						N	x			

(*Continued*)

Table 2. (*Continued*)

No.[1]	VOC[2]	Retention time[3]	H healthy non smoker % pos.[4]	H COPD ex-smoker % pos.[5]	H healthy smoker % pos.[6]	H COPD smoker % pos.[7]	M healthy non smoker % pos.[8]	M COPD ex-smoker % pos.[9]	M healthy smoker % pos.[10]	M COPD smoker % pos.[11]	% pos samples in room air[12]	Tenax decomp.[13]	Low abundance[14]	Smokers only[15]	Smoking related[16]	Room air related[17]	Corr. breath/room air samples[18]	Room air site specific[19]	Breath site specific[20]	Gender specific	Not reproducible	Range (ppb)[21]	COPD related (univariate) all[22]	COPD related (univariate H)[23]	COPD related (univariate M)[24]	Included into analysis[25]	Included in non/ex-smoker model[26]	Included in smoker model[27]	COPD related site/ smoking status[28]
78	Decanitrile	22.7	96	96	100	100	100	89	93	100	62															x			x
129	Undecanal	22.7	100	100	100	100	100	93	100	100	93					x													
79	2-Hydroxy benzoic acid ethyl ester	22.8	74	68	67	48	12	0	0	0	3								H							x			x
130	Decanoic acid methyl ester	22.9	30	32	38	48	0	4	0	18	1								H							x			
131	Anethole	23.1	70	57	58	56	24	19	37	45	8															x			
132	3-p-Menthene	23.2	70	75	67	74	72	52	47	82	21															x			
133	Menthyl acetate	23.2	22	21	29	52	32	22	23	9	0															x			
134	Citronellol acetate	23.7	33	39	29	44	32	33	23	18	0															x			
80	Tetradecane	23.9	100	100	100	100	100	89	90	100	85						M					0.1–1				x			
81	**Indole**	24.0	100	100	96	100	96	67	87	64	7												N/S	S	N/S	x			x
135	Nerol acetate	24.5	96	93	92	96	100	93	90	100	12															x			x

Notes. 1: random numbering of VOCs. 2: VOC name, *normal*: identification by NIST database (compound with highest probability), *bold*: considered as COPD related in this study. 3: retention time. 4–11: percentage of positive samples in the 4 groups in Hannover and Marburg. 12: percentage positive samples in room air measurements. 13: considered *as Tenax® decomposition* related: *bold*, described in [32]. 14: *low abundant*: less than 25 positive subjects per site or less than 12 positive non/ex-smoking subjects. 15: only found in active smokers. 16: considered as smoking related: *bold*: significantly ($p < 0.001$) higher levels in smokers and significantly correlated to exhaled CO ($r > 0.6$; $p < 0.001$). 17: marked, if median breath to room air ratio <1.1 in healthy subjects both in Marburg and Hannover and if the percentage of room air positive samples >30%. 18: H: correlations between breath and room air samples only in Hannover, M: only in Marburg, A: (all) in data from both sites (visual control of correlation plots). 19: letter indicates significant difference between sites, M: with higher levels in Marburg, H with higher levels in Hannover. 20: *bold*: difference between sites at 1 site, H: higher levels in Hannover, M: higher levels in Marburg. 21: range of concentrations for selected VOCs in our study. Data based on standard curves using pure VOCs for compound identification. For compound No.15, 20 and 61 no standard curve available. 22–24: *normal*: indicates univariate difference between COPD patients and controls in either smokers (S), non/ex-smokers (N) or both groups N/S with $p < 0.001$, *normal*: more than 2 × more positive cases at 1 site, Hannover (23), Marburg (24) or in both sites together (22) a difference with $p < 0.005$ was observed (indicated in *bold*). 25: VOCs included into the analysis (reason for exclusion see discussion). 26: indicates inclusion into the model to discriminate COPD patients from controls in non/ex-smokers and 27: in active smokers. 28: indicates those VOC that show the same difference between COPD patients and controls in smokers and non/ex-smokers. 28: *bold* indicates that the VOC shows the same difference between COPD patients and controls in all subjects, within each site and both in smokers and non/ex-smokers.
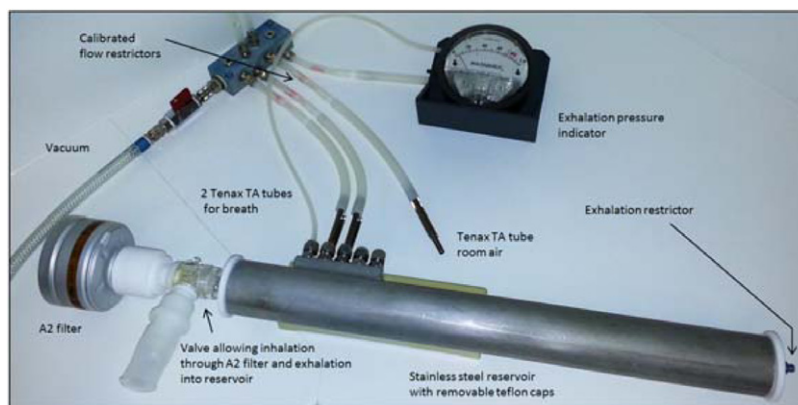
Figure 1. Sampling device. Subjects inhaled through the A2 filter on the left and exhaled into the stainless steel reservoir and flow control (resistor at the right hand end of the reservoir and pressure indicator for visual control in the background). From this reservoir 2 Tenax® TA were continuously loaded over a 5 min period (flow through the tubes controlled at 500 ml min^{-1}).
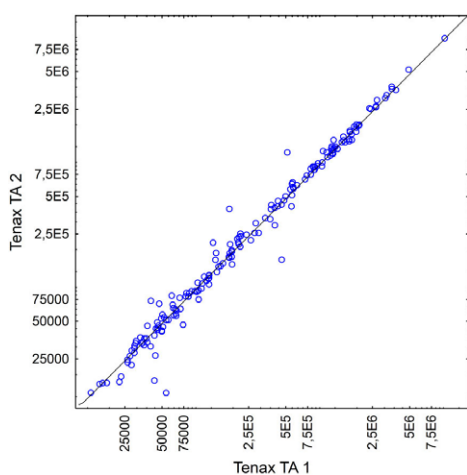

Figure 2. Correlation for the level of toluene between two Tenax® TA tubes loaded simultaneously using the sampler shown in figure 1. Graph also shows the line of identity.
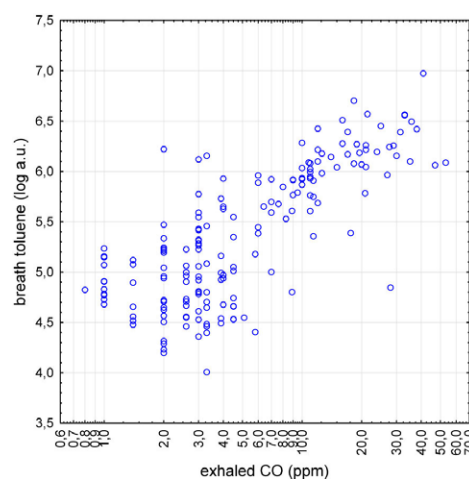
Figure 3. Correlation between the level of toluene and CO in exhaled breath.

and carboxyhemoglobin (COHb) as well as lower NO values compared to non- and ex-smokers.

Reproducibility of measurements, transport BTX levels

We compared the VOC levels from two independently loaded Tenax® tubes to search and to correct for outliers. VOC levels were generally highly reproducible (figure 2), but fluctuations in retention time sometimes led to an insufficient peak integration that had to be manually corrected. Heptane partly overlapped with acetic acid in the chromatogram and m/z 43 showed to be the main fragmentation ion for both compounds. However targeting on alternative quantifier did not improve the quality of automatically generated results sufficiently and therefore heptane was excluded from the analysis.

It is important to note, that the level of benzene, toluene, xylene (BTX) aromatic hydrocarbons was not significantly elevated in samples of healthy non-smokers from Marburg compared to non-smokers from Hannover. Analyzing all non/ex-smokers we even found slightly higher values in the Hannover samples, for toluene this

difference was significant ($p < 0.01$). In room air samples we also observed slightly higher levels of BTX compounds in Hannover, for the level of benzene this difference was statistically significant ($p < 0.05$). On the other hand the effect of active smoking on the level of BTX was clearly detectable and these VOCs were related to the level of CO in the breath of smokers (table 2, figures 3 and 4).

VOCs in exhaled breath

134 VOCs basically representing the major peaks observed in the chromatograms of the different subject groups or of room air samples were randomly selected for the analysis (table 2). Depending on the subject group the selection corresponds to approximately 80–90% of peaks that would have been included into the automated analysis of the Chemstation software. Table 2 summarizes the available information concerning these VOCs, including retention time, percentage of positive findings in the different subject groups and in room air samples. Eight of the VOCs in the list were Tenax® related (e.g. acetophenone, benzoic acid, benzaldehyde) and therefore excluded from the further analysis. Phenol, which can also be related

to Tenax® matrix decomposition was increased in exhaled breath samples compared to room air samples and therefore not excluded. Fourteen VOCs were considered as low abundant, because they were detected in less than 25 subjects per site, and excluded from the analysis except for 2, which were considered as smoking related. Thirty-two VOCs were also found in significant amounts in room air samples at both sites (median ratio breath/room air <1.1), of which we excluded 18 from the analysis, as these were not or only in few exhaled breath samples observed in higher concentrations compared to room air. Typical chromatograms of the different subject groups of each study site as well as room samples are provided in the online supplement.

Smoking related VOCs
Eight VOCs were only detected in active smokers, additional 46 (including phenol) were considered as smoking related, because higher levels were found in actively smoking subjects. Active smokers could be clearly discriminated by higher values for combustion products like furans as well as BTX aromatic hydrocarbons. Table 2 provides a list with those VOCs that were significantly elevated in active smokers and also shows that most of these correlate with the CO or COHb values that were obtained at the time of VOC breath collection. The relationship between these independent measurements indicates a valid breath sampling methodology and VOC analysis. As higher numbers of pack-years (PY) are related to the amount of active smoking at the time of breath sampling, there were also weak correlations between the smoking related VOCs and PY with correlation coefficients <0.5.

To test our LDA approach we assessed whether 'smoking status' would be correctly classified (table 3). For this analysis we included the 101 VOCs that are listed in table 2. Of the 10 VOCs included into the model most were considered as smoking related VOCs. Interestingly, 7 of the 8 misclassified smokers had CO levels below 6 ppm and overall the group of the 8 misclassified smokers had a significantly ($p < 0.003$) lower level of exhaled CO compared to the 78 correctly classified smokers.

Comparison between sites
Within the list of 134 VOCs used for exhaled breath analysis we found significant differences ($p < 0.001$) for room air between sites for e.g. ethanol, isoprene, acetone, room cleaning related compounds or menthol (table 2). For hexanoic acid (increased in Marburg) and for 2-methyl benzofurane (increased in Hannover) these differences were also observed in breath of healthy non-smoking controls, though limited to a subset of subjects. Some lower abundant VOCs that were not included into the overall analysis were also site specific (table 2).

Table 2 shows the 5 breath VOCs that were found to be significantly higher in all groups between sites ($p < 0.001$) and those additional 12 VOCs that had

Table 3. Classification for smoking (CV by leave-one-out (LOO) method).

Smoke			Predict. group 0	Predict. group 1	Total
Training	n	0	102	2	104
		1	8	78	86
	%	0	98.1	1.9	100.0
		1	9.3	90.7	100.0
Cross-valid	n	0	102	2	104
		1	10	76	86
	%	0	98.1	1.9	100.0
		1	11.6	88.4	100.0

Table 4. Classification for site (CV by leave-one-out (LOO) method).

Site			Predict. group 0	Predict. group 1	Total
Training	n	0	104	1	105
		1	0	85	85
	%	0	99.0	1.0	100.0
		1	0.0	100.0	100.0
Cross-valid.	n	0	104	1	105
		1	0	85	85
	%	0	99.0	1.0	100.0
		1	0.0	100.0	100.0

more than twice the number of positive cases at a site compared to the other.

Using LDA and including the 101 selected VOCs (table 2) into the analysis almost resulted in a perfect separation of the sites (table 4). 19 VOCs were included into this model, for 3 VOCs we found evidence for a room air relationship. Excluding these from the analysis did not change the result. Both the univariate and the multivariate data indicate that independent from environmental factors differences in breath VOCs exist.

COPD related breath VOC
For the search of COPD related VOCs we used a multivariate LDA as well as the univariate comparison between healthy subjects and patients with COPD. Due to the known profound effect of smoking, both approaches were performed separately for non/ex-smokers and for active smokers.

For the LDA analysis we included all 190 subjects and used the 101 VOCs indicated (index 25) in table 2, which were also used for the discrimination of smokers and non-smokers and for site described above. The result of this analysis is displayed in tables 5 and 6.

8 VOCs were included into the model to separate the non/ex-smokers, and 4 different VOCs were included to separate the actively smoking groups, however, 6/8 VOCs included into the model to discriminate COPD patients from controls in non/ex-smokers and 3/4 VOCs included into the model for active smokers showed similar differences in the respective other group

Table 5. Classification for diagnosis in the group of non/ex-smokers.

Diagnosis non/ex-smokers			Predict. group		
			0	1	Total
Training	n	0	48	4	52
		1	7	45	52
	%	0	92.3	7.7	100.0
		1	13.5	86.5	100.0
Cross-valid.	n	0	45	7	52
		1	8	44	52
	%	0	86.5	13.5	100.0
		1	15.4	84.6	100.0

Table 6. Classification for diagnosis in the group of actively smoking subjects.

Diagnosis smokers			Predict. group		
			0	1	Total
Training	n	0	39	10	49
		1	5	32	37
	%	0	79.6	20.4	100.0
		1	13.5	86.5	100.0
Cross-valid.	n	0	37	12	49
		1	7	30	37
	%	0	75.5	24.5	100.0
		1	18.9	81.1	100.0

(marked bold in the respective columns of table 2). This could be interpreted as additional evidence for a relationship of these VOCs to COPD.

We also performed the analysis using both sites as training sets and the respective other site as an independent validation set. The training datasets of the separate sites lead to comparable levels of discrimination between healthy subjects and COPD patients, with a lower rate of correct detection in Hannover compared to Marburg. The respective validation of the different models derived from the site specific analysis, however, yielded unacceptable high error rates. Limiting the analysis to those VOCs considered as relevant from the whole group analysis did not improve the error rates substantially.

Therefore we also performed the univariate analysis between healthy and COPD patients separately for active smokers and non/ex-smokers. The bold values in table 2 indicate the VOCs with significant ($p < 0.005$) differences between groups separately for the whole group and for the two sites. If in the whole group or in one of the sites such a clear difference was found, than it is also indicated, that a comparable difference with $p < 0.02$ was also observed at the respective other site or within the whole group. The table shows, that in Marburg we found a much larger number of VOCs with differences between healthy subjects and COPD patients and that these were predominantly found in smoking subjects. In Hannover only a few VOCs with significant differences were observed, and these were mainly in the non/ex-smokers.

Figure 4 shows the 7 VOCs that were significantly ($p < 0.05$) different in the whole group and showed the same kind of difference between healthy subjects and COPD patients both in non/ex-smokers and smokers. Interestingly, only vinyl acetate and 1-ethyl-3-methyl benzene are among the VOCs that were included into the multivariate models.

Overall we found evidence for a relationship to COPD for 14 VOCs. None of the 14 VOCs showed a correlation with age, neither when including all subjects, nor within the 4 groups of subjects separately. We also found no relationship with the FEV/FVC ratio.

Discussion

In this study we used continuous breath sampling from an inert reservoir and analyzed selected target ion identified VOCs for their relationship to stable COPD. Based on (a) the good correlation between the data of two simultaneous collected adsorption tubes, (b) the correlation between smoking related VOCs and the independently assessed levels of exhaled CO, and (c) the ability to discriminate between smokers and non-smokers with a low error rate we assume a sufficient validity of our data. From the multi- and the univariate analysis we consider 14 of the 134 selected VOCs to be COPD related. The evidence is either derived from the inclusion into the respective LDA models or from the fact that a comparable difference between controls and COPD patients was not only observed in the analysis of all subjects, but also separately for the subjects of each site and within the groups of smokers and non/ex-smokers. In addition, site specific differences were detected for a number of VOCs. These extend beyond environmental and cleaning related compounds, therefore we consider this to be an important result that needs to be taken into account in future breath research trials. In line with others we found that smoking has a profound effect on the composition of breath VOCs [26, 33], emphasizing the need to adequately control for smoking status especially in studies investigating smoking related diseases such as COPD [12].

Smoking

The effect smoking, on the other hand, can be used to validate breath VOC data and the respective statistical analysis procedures. As expected, we were able to discriminate between non/ex-smokers and smokers in the multivariate analysis. Interestingly, the misclassifications were almost all smokers that were considered as non-smokers. They might have had longer periods than the recommended 2 h of non-smoking or smoked fewer cigarettes prior to their visit, as they had significantly lower CO levels compared to the rest of the smokers. The independently assessed breath CO levels also support the validity of our GC/MS data, as they showed a good correlation with most smoking related VOCs. For some of these, like BTX hydrocarbons and 2,5-dimethyl-furane,

comparable effects of active smoking have also been reported by Alonso and coworkers [33].

Site

Both for room air VOCs and for breath VOCs we found differences between the two sites Marburg and Hannover. The room air differences could be in part attributed to differences in overall cleaning procedures. Some of the site specific VOCs would also suggest differences in the number of staff in the vicinity of breath sampling or in room air exchange rates between the sites. With respect to the differences in BTX aromatic hydrocarbons both in room air and in breath we would assume that the size and location of the sites played a role. Marburg is a small city in a rural area 20 km away from a major highway with just 80 000 inhabitants, while Hannover has about 500 000 inhabitants, and is linked to several major highways. Our results did not provide evidence for a contamination of Marburg samples during road shipment by courier. Differences in VOC pattern between two sampling sites in China and Latvia have also been reported before [34]. Besides the differences in room air we also found some very pronounced differences in breath VOCs, which could be attributed to differences in breathing valves and/ or their respective cleaning procedures. Examples for these differences are provided in the online supplement (figures S1–S3). These extreme site specific VOCs were not included into the analysis and the 134 VOCs listed in table 2. Excluding all room air or suspected cleaning related VOCs from the analysis still resulted in almost complete discrimination between subjects of both sites, indicating that additional site specific differences in exhaled breath exist. These could be environmental or lifestyle-related and have to be taken into account in multicenter studies.

COPD

To identify COPD related VOCs we compared the VOC patterns between healthy subjects and COPD patients separately in the group of non/ex-smokers and in smokers. Merged evidence from the univariate and multivariate analysis resulted in a list of 14 VOCs with a potential relationship to COPD. We considered the fact that a VOC was included into the multivariate model of a group and at the same time showed a comparable difference between controls and COPD in the respective other group as evidence. If the VOC also showed the same difference between controls and COPD patients separately at each site, the evidence was considered as strong (figure 4).

Analysis of the data from the separate sites as training and test datasets did not lead to models with sufficient discriminative power. Even focusing the analysis on the 14 potentially COPD related VOCs lead to unacceptable high error rates. This could be due to absolute differences in the levels of these VOCs between sites or due to the variability of the data (see individual data for selected VOCs in figure 4). There were also some differences between the patients groups recruited at the study sites with respect to gender distribution and lung function, but currently their effect on the results are difficult to interpret.

The results of the multivariate analysis are affected by the number and kind of variables included. Adding a VOC or leaving one out can have an effect on the VOCs included into the models to discriminate COPD patients from controls. Therefore, our pre-selection of 101 VOCs for the analysis is likely to affect our results, but more in a sense that we might have missed discriminative VOCs. However, with the currently existing differences in sampling procedures, adsorption material, analysis technology, and statistical evaluation methods between study groups, a complete set of COPD related VOCs cannot be expected from a single study. But every study can contribute to the list of potentially COPD related VOCs that then can be tested in larger population- and multi-center studies for their clinical relevance. This includes to test how stable these VOC markers are over time and how they respond to changes in e.g. airway inflammation, treatment, exacerbation rate and potentially to other disease related outcomes.

Comparison with other COPD studies

There are a number of studies reporting differences between patients with COPD and healthy control subjects. The results of studies using sensor arrays or eNose technology [9, 14, 23] cannot be compared with our data.

This is also true for studies using very sensitive ion mobility spectrometry (IMS), as these generally do not identify the discriminating peaks. One example is the recently published study by Besa and colleagues reporting differences in breath VOC patterns for 45 COPD patients, 23 healthy smokers and 28 healthy non-smokers [8]. Interestingly, however, that there appeared to be a smaller percentage of smoking related VOCs among the discriminating peaks as compared to our study. Using a real time MS, Sinues and coworkers found a similar level of discrimination between controls and COPD patients [17]. One of their major discriminating VOCs was most likely indole, which also showed lower levels in the COPD patients in our study.

In 2012, Phillips and coworkers performed a large COPD study including 119 stable COPD patients (34% active smokers) and 63 healthy controls (10% active smokers) [12]. The major difference compared to our study was that just 130 ml samples were collected and loaded onto carbon based adsorption tube material. Similar to the study by Besa *et al*, approximately half the patients were GOLD 3 and 4 and therefore more severely ill as compared to the patients included into our study. Despite the differences in sampling and analysis, Phillips *et al* also found evidence for BTX aromatic hydrocarbons, acetic acid, and phenol to be COPD related.

Basanta and coworkers compared 39 COPD patients (31% active smokers) and 32 healthy subjects
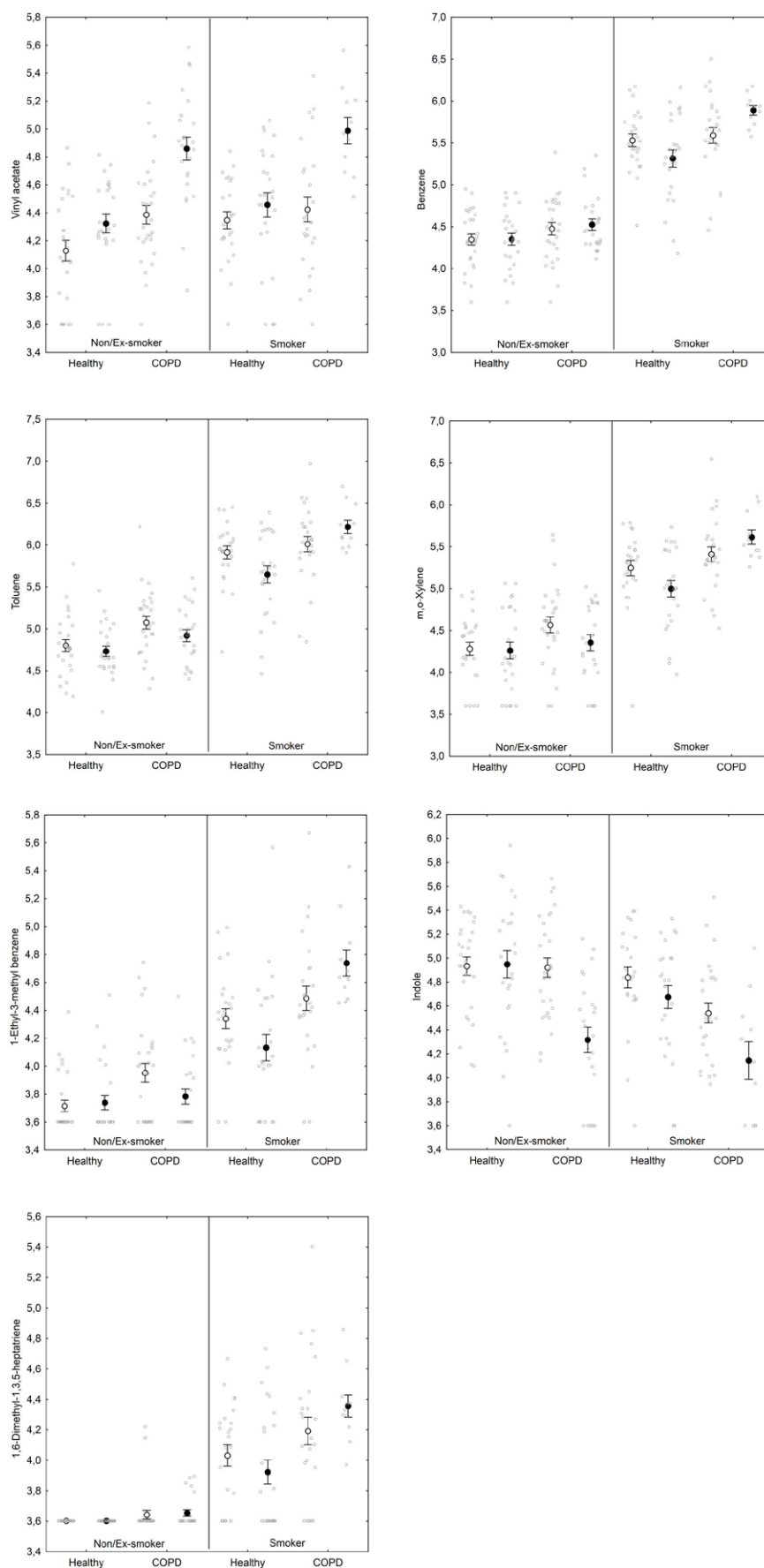
Figure 4. Seven VOCs showing a difference between controls and COPD patients in all subjects, in subjects of each site (open circles Hannover, closed circles: Marburg) and in non/ex-smokers (left) and smokers (right). Mean and SE of log transformed data. Individual data is displayed in grey. Identification of VOCs by target ions, retention time, Nist database or reference compound (table 2).

(31% active smokers) [11] using a special sampling device that enabled a direct loading of late expiratory breath onto Tenax® TA/Carbotrap adsorption tubes. Comparable to our study 3 l of breath were analyzed, but the use of GC-time of flight (ToF)-MS 487 VOCs were identified. The COPD subject, which were predominantly GOLD II and III could also be discriminated from the healthy controls, however, there is no overlap between the VOCs used for model classification by Basanta *et al* and those used in our study. Hexanal, which was excluded from our analysis due to high levels in room air, was the only substance that overlaps with the markers listed by Phillips *et al* [12]. Nonanal and decanal, which were also considered as discriminative between healthy subjects and COPD by Basanta were excluded from our analysis, as both are considered Tenax® TA artifacts [32]. For these VOCs we observed a clear correlation between breath samples and room air samples, potentially indicating varying levels of trap Tenax® TA decomposition.

Van Berkel and coworkers recruited 50 COPD patients (76% active smokers) and 29 controls (31% active smokers) into their study. In addition they included 16 COPD patients (12% active smokers) and 16 controls (19% active smokers) from a different hospital for validation. Breath sampling was performed by filling Tedlar bags with 5 l of breath which were transferred to adsorption tubes comparable to those used by Phillips *et al* as described above. Analysis was performed by TD-GC-TOF-MS. Again, hexanal was among the potentially discriminating VOCs, which in opposite to our results and e.g. data by Basanta *et al* all showed lower levels and lower frequencies in the COPD patients. Interestingly, the training model by van Berkel *et al* with 6 VOCs (not including hexanal) showed a nearly optimal performance in the validation dataset from a different hospital, indicating that their 6 VOC classifiers are independent of site and smoking.

Selection and analysis of VOCs

We selected 101 VOCs for the final analysis. The main reason to exclude a VOC from this list was that the VOC (1) was suspected to be a Tenax® TA decomposition product, (2) low abundance or (3) room air related unless we found evidence for a relationship to smoking. Apart from these rules VOC 14 and 116 (table 2) remained in the list as we found evidence for a relationship to COPD from the univariate analysis and VOC 56 and 125, as a correlation between breath and room air was only found at one site. The exclusion of VOCs from the analysis bears the risk of losing potential information and missing discriminating VOCs. On the other hand, we feel that available information about a VOC e.g. its relationship to room air or to cleaning procedures should be taken into account if possible. In addition, limiting the number of variables in studies with limited number of patients provides a more robust basis for the statistical analysis and reduces the risk of over fitting.

Limitations of our study

Despite the fact that we performed one of the largest exhaled breath studies to find COPD related VOCs by including 190 patients and controls, studies with much larger patient numbers are required to provide data for robust statistical analysis. However, these require a multicenter approach with standardized sampling and analysis procedures that enables to build a large disease related database. Efforts for a standardization of procedures are under way and a recently introduced novel sampling device (Owlstone, UK [35]) could potentially support this.

Our study provided cross-sectional data for COPD related VOCs, which cannot be used to predict any COPD outcomes or help to monitor the disease. The clinical value of the potentially COPD related VOCs will need to be evaluated in longitudinal trials. Being well aware that COPD is a multifactorial, very complex disease the suggested markers of our and other comparable studies are also unlikely to improve the diagnosis of COPD at the current stage.

We did not include external standards into our analysis. Therefore we have to deal with a certain variability of our data due to changes in loading capacity of Tenax® tubes or due to instrument drift. As our patients were recruited in random order, we consider it unlikely that we introduced a major bias into our data due to the lack of such a standard. However, it might have affected the analysis by increasing the overall variability of the data.

Using Tenax® TA as adsorption material has advantages, e.g. useful for high moisture content samples like breath, low desorption temperature with less risk to produce unwanted artifacts, but also some drawbacks. There are a number of known decomposition products, some of which, like benzaldehyde have been suggested by others for data normalization [12, 26]. In addition, nonanal and decanal, which have been suggested as potentially COPD related, are difficult to evaluate correctly when Tenax® TA is used as adsorption material.

In this study we did not use all the available information provided by the chromatograms to discriminate between groups. Instead we identify 134 VOCs by target ions and by NIST database. This approach is likely to miss information especially of lower abundant or lower level VOCs. On the other hand the identification by target ions allows the separation of VOCs with overlapping retention times which is not possible when using the TIC as outcome parameter. Using a growing database with retention times and target ions is also an economic way to deal with the large amount of data, avoiding pre-processing like alignment of chromatograms or baseline subtraction, which could have unexpected effects on the data if performed unsupervised.

Conclusion

In summary, we identified 10 novel breath VOCs that appear to be related to COPD using a different

approach in breath sampling compared to others and by using a limited number of target ion identified VOCs. We confirm the profound effect of smoking in our study and show that so far unexplained site specific differences exist, that need to be taken into account when interpreting results from multicenter studies.

Acknowledgment

References

[1] van de Kant K D, van der Sande L J, Jobsis Q, van Schayck O C and Dompeling E 2012 Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review *Respir. Res.* **13** 117

[2] Fens N, van der Schee M P, Brinkman P and Sterk P J 2013 Exhaled breath analysis by electronic nose in airways disease. Established issues and key questions *Clin. Exp. Allergy* **43** 705–15

[3] de Lacy C B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D and Ratcliffe N M 2014 A review of the volatiles from the healthy human body *J. Breath Res.* **8** 014001

[4] Bikov A, Lazar Z and Horvath I 2015 Established methodological issues in electronic nose research: how far are we from using these instruments in clinical settings of breath analysis? *J. Breath Res.* **9** 034001

[5] Sethi S, Nanda R and Chakraborty T 2013 Clinical application of volatile organic compound analysis for detecting infectious diseases *Clin. Microbiol. Rev.* **26** 462–75

[6] Zhu J, Bean H D, Wargo M J, Leclair L W and Hill J E 2013 Detecting bacterial lung infections: *in vivo* evaluation of in vitro volatile fingerprints *J. Breath Res.* **7** 016003

[7] Sibila O *et al* 2014 Identification of airway bacterial colonization by an electronic nose in chronic obstructive pulmonary disease *Respir. Med.* **108** 1608–14

[8] Besa V, Teschler H, Kurth I, Khan A M, Zarogoulidis P, Baumbach J I, Sommerwerck U, Freitag L and Darwiche K 2015 Exhaled volatile organic compounds discriminate patients with chronic obstructive pulmonary disease from healthy subjects *Int. J. Chron. Obstruct. Pulmon. Dis.* **10** 399–406

[9] Fens N, Zwinderman A H, van der Schee M P, de Nijs S B, Dijkers E, Roldaan A C, Cheung D, Bel E H and Sterk P J 2009 Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma *Am. J. Respir. Crit. Care Med.* **180** 1076–82

[10] Incalzi R A, Pennazza G, Scarlata S, Santonico M, Petriaggi M, Chiurco D, Pedone C and D'Amico A 2012 Reproducibility and respiratory function correlates of exhaled breath fingerprint in chronic obstructive pulmonary disease *PLoS ONE* **7** e45396

[11] Basanta M, Ibrahim B, Dockry R, Douce D, Morris M, Singh D, Woodcock A and Fowler S J 2012 Exhaled volatile organic compounds for phenotyping chronic obstructive pulmonary disease: a cross-sectional study *Respir. Res.* **13** 72

[12] Phillips C O, Syed Y, Parthalain N M, Zwiggelaar R, Claypole T C and Lewis K E 2012 Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls *J. Breath Res.* **6** 036003

[13] Fens N *et al* 2011 Exhaled air molecular profiling in relation to inflammatory subtype and activity in COPD *Eur. Respir. J.* **38** 1301–9

[14] Hattesohl A D, Jorres R A, Dressel H, Schmid S, Vogelmeier C, Greulich T, Noeske S, Bals R and Koczulla A R 2011 Discrimination between COPD patients with and without alpha 1-antitrypsin deficiency using an electronic nose *Respirology* **16** 1258–64

[15] Phillips C, Mac P N, Syed Y, Deganello D, Claypole T and Lewis K 2014 Short-term intra-subject variation in exhaled volatile organic compounds (VOCs) in COPD patients and healthy controls and its effect on disease classification *Metabolites* **4** 300–18

[16] Schivo M, Seichter F, Aksenov A A, Pasamontes A, Peirano D J, Mizaikoff B, Kenyon N J and Davis C E 2013 A mobile instrumentation platform to distinguish airway disorders *J. Breath Res.* **7** 017113

[17] Martinez-Lozano S P, Meier L, Berchtold C, Ivanov M, Sievi N, Camen G, Kohler M and Zenobi R 2014 Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease *Respiration* **87** 301–10

[18] Cristescu S M, Gietema H A, Blanchet L, Kruitwagen C L, Munnik P, van Klaveren R J, Lammers J W, Buydens L, Harren F J and Zanen P 2011 Screening for emphysema via exhaled volatile organic compounds *J. Breath Res.* **5** 046009

[19] Basanta M, Jarvis R M, Xu Y, Blackburn G, Tal-Singer R, Woodcock A, Singh D, Goodacre R, Thomas C L and Fowler S J 2010 Non-invasive metabolomic analysis of breath using differential mobility spectrometry in patients with chronic obstructive pulmonary disease and healthy smokers *Analyst* **135** 315–20

[20] Dragonieri S, Annema J T, Schot R, van der Schee M P, Spanevello A, Carratu P, Resta O, Rabe K F and Sterk P J 2009 An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD *Lung Cancer* **64** 166–70

[21] Beauchamp J D and Pleil J D 2013 Simply breath-taking? Developing a strategy for consistent breath sampling *J. Breath Res.* **7** 042001

[22] Koczulla A R *et al* 2011 Smelling diseases? A short review on electronic noses *Pneumologie* **65** 401–5

[23] Fens N, Roldaan A C, van der Schee M P, Boksem R J, Zwinderman A H, Bel E H and Sterk P J 2011 External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease *Clin. Exp. Allergy* **41** 1371–8

[24] Van Berkel J J, Dallinga J W, Moller G M, Godschalk R W, Moonen E J, Wouters E F and Van Schooten F J 2010 A profile of volatile organic compounds in breath discriminates COPD patients from controls *Respir. Med.* **104** 557–63

[25] Bos L D, Weda H, Wang Y, Knobel H H, Nijsen T M, Vink T J, Zwinderman A H, Sterk P J and Schultz M J 2014 Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome *Eur. Respir. J.* **44** 188–97

[26] Filipiak W *et al* 2012 Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants *J. Breath Res.* **6** 036008

[27] Kirk A T, Allers M, Cochems P, Langejuergen J and Zimmermann S 2013 A compact high resolution ion mobility spectrometer for fast trace gas analysis *Analyst* **138** 5200–7

[28] Langejuergen J, Wagner C, Beutel S, Hopmeier T, Scheper T and Zimmermann S 2015 Non-invasive monitoring of bacterial growth and auto-induced protein production in a bioreactor with a closed-loop GC-IMS *Int. J. Ion Mobil. Spectrom.* **18** 9–15

[29] Allers M, Langejuergen J, Gaida A, Holz O, Schuchardt S, Hohlfeld J and Zimmermann S 2015 Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using a closed-loop GC-IMS *Int. Association of Breath Research IABR—Summit 2015*

[30] Langejuergen J, Gaida A, Holz O, Schuchardt S, Hohlfeld J and Zimmermann S 2014 Ultra-sensitive atmospheric pressure chemical ionization mass spectrometry for breath analysis *Int. Association of Breath Research IABR—Summit 2014*

[31] Allers M, Langejuergen J, Gaida A, Holz O, Schuchardt S, Hohlfeld J and Zimmermann S 2016 Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using closed gas loop GC-IMS and GC-APCI-MS *J. Breath Res.* **10** 026004

[32] Lee J H, Batterman S A, Jia C and Chernyak S 2006 Ozone artifacts and carbonyl measurements using Tenax GR, Tenax TA, Carbopack B and Carbopack X adsorbents *J. Air. Waste Manage. Assoc.* **56** 1503–17

[33] Alonso M, Castellanos M and Sanchez J M 2010 Evaluation of potential breath biomarkers for active smoking: assessment of smoking habits *Anal. Bioanal. Chem.* **396** 2987–95

[34] Amal H, Leja M, Broza Y Y, Tisch U, Funka K, Liepniece-Karele I, Skapars R, Xu Z Q, Liu H and Haick H 2013 Geographical variation in the exhaled volatile organic compounds *J. Breath Res.* **7** 047102

[35] Kitchen S, Edge A, Smith R, Thomas P, Fowler S, Siddiqui S and van der Schee M P 2015 Breathe free: open source development of a breath sampler by a consortium of breath researchers *Eur. Resp. J.* **46** (Suppl. 59) (doi: 10.1183/13993003.congress-2015.PA3987)