Journal of Breath Research

PAPER

Detection of *Staphylococcus aureus* in cystic fibrosis patients using breath VOC profiles

RECEIVED
14 July 2016

REVISED
24 October 2016

ACCEPTED FOR PUBLICATION
1 November 2016

PUBLISHED
30 November 2016

A H Neerincx[1,3,6], B P Geurts[2,6], J van Loon[3], V Tiemes[3], J J Jansen[2], F J M Harren[1], L A J Kluijtmans[4], P J F M Merkus[3,5], S M Cristescu[1], L M C Buydens[2] and R A Wevers[4]

[1] Department of Molecular and Laser Physics, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands
[2] Department of Analytical Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands
[3] Department of Paediatrics, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, The Netherlands
[4] Department of Laboratory Medicine, Translational Metabolic Laboratory, Radboud University Medical Center, Nijmegen, The Netherlands
[5] Department of Paediatrics, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands
[6] Authors contributed equally to this work.

E-mail: s.cristescu@science.ru.nl

Keywords: *Staphylococcus aureus*, breath analysis, volatile organic compound (VOC) profiles, GC-MS, sparse partial least squares discriminant analysis (s-PLS-DA), cystic fibrosis

Abstract

Staphylococcus aureus (*S. aureus*) is a common bacterium infecting children with cystic fibrosis (CF). Since current detection methods are difficult to perform in children, there is need for an alternative. This proof of concept study investigates whether breath profiles can discriminate between *S. aureus* infected and non-infected CF patients based on volatile organic compounds (VOCs). We collected exhaled breath of CF patients with and without *S. aureus* airways infections in which VOCs were identified using gas chromatography–mass spectrometry. We classified these VOC profiles with sparse partial least squares discriminant analysis. Multivariate breath VOC profiles discriminated infected from non-infected CF patients with high sensitivity (100%) and specificity (80%). We identified the nine compounds most important for this discrimination. We successfully detected *S. aureus* infection in CF patients, using breath VOC profiles. Nine highlighted compounds can be used as a focus point in further biomarker identification research. The results show considerable potential for non-invasive diagnosis of airway infections.

1. Introduction

Respiratory infections are the main problem in patients with cystic fibrosis (CF), since they result in progressive loss of lung function and eventually death [1]. *Staphylococcus aureus* (*S. aureus*) is one of the most common bacteria infecting children with CF. On average, 68% of patients with CF are infected with *S. aureus* (including people with methicillin-resistant *S. aureus* (MRSA) infection) [2]. The prevalence of *S. aureus* infection peaks at almost 80% in six- to ten-year old patients [3]. Early detection of respiratory infections is essential for adequate treatment, but currently available detection methods are difficult to perform in children, time consuming, invasive (bronchoalveolar lavage), non-specific (chest x-ray) or insensitive and unpleasant (sputum induction) [3–6]. For this reason, a non-invasive, easy, fast and accurate method to detect respiratory infections is needed.

One alternative approach could be the analysis of exhaled breath. Up till now, potential markers specific for *S. aureus* detection *in vitro* are reported to be isovaleric acid and methylbutanal [7, 8]. Other compounds elevated in *S. aureus* headspace are ammonia, methanol, acetaldehyde, ethanol, propanol and pentanal [9], as well as butanol and acetic acid [10], acetone [10, 11], 2-methylbutanal, benzaldehyde, 2,3-butanedione, 1-methyl-4-(1-methylethenyl)cyclohexane, dimethyldisulfide and dimethyltrisulfide [11], 3-methyl-butanal [11, 12], and 3-methyl-1-butanol, butanoic acid and 3-methyl-butanoic acid [12]. However, these compounds are also found in the headspace of other pathogens [7–9, 12–16] and are therefore not specific for *S. aureus*. *In vivo*, acute *S. aureus* infection was detected in a bacterial murine lung infection model, in an untargeted metabolomics study using secondary electrospray ionization-mass spectrometry (SESI-MS) breath profiles [17]. Furthermore, *S. aureus*-derived

Table 1. Patient characteristics by *S. aureus* infection status.

Patient number	Gender	Age (years)	*S. aureus* infection	Other infections
1	Female	6	Yes	—
2	Female	7	Yes	—
3	Female	9	Yes	—
4	Female	13	Yes	*Serratia marcescens*
4	Female	13	Yes	*Pseudomonas aeruginosa*
5	Female	21	Yes	*Pseudomonas aeruginosa* and *yeast*
6	Female	23	Yes	*Haemophilus influenzae*
7	Male	8	Yes	—
8	Male	9	Yes	*Haemophilus influenzae*
9	Male	11	Yes	—
10	Male	12	Yes	*Haemophilus influenzae*
11	Male	16	Yes	—
12	Male	19	Yes	—
13	Male	54	Yes	*Haemophilus influenzae*
14	Female	7	No	—
15	Female	7	No	—
16	Male	19	No	*Stenotrophomonas maltophilia* and *Aspergillus fumigatus*
17	Male	7	No	—
18	Male	7	No	—

metabolites could be detected in ventilator-associated pneumonia in intensive care patients, in a recent prospective pilot study [18]. However, to our knowledge, *in vivo* studies for detection of *S. aureus* in exhaled breath of CF patients have not been performed yet.

Therefore, the aim of this proof of concept study was to investigate whether we can discriminate between *S. aureus* infected and non-infected CF patients based on volatile organic compounds (VOCs) in their exhaled breath. For this purpose, we used gas chromatography–mass spectrometry (GC-MS) for breath VOC analysis, in combination with sparse partial least squares discriminant analysis (s-PLS-DA) for multivariate discrimination between the different patient groups.

2. Materials and methods

2.1. Subjects

CF patients (⩾6 years old) were recruited between January 8 and July 16, 2015, from the outpatient clinic of Radboud University Medical Center—Amalia Children's Hospital, Nijmegen, the Netherlands. This study was approved by the local medical ethics committee (CMO—Nijmegen/Arnhem) which waived written informed consent. Oral consent was obtained from both participants aged 12 years and older, and from parents of children aged 6–18 years.

Non-smoking CF patients with and without *S. aureus* infection were included in this study (table 1). Infection was defined as culturing *S. aureus* in sputum (or cough swab sample for children unable to expectorate), on the day of breath sampling or in ⩾50% of the samples (minimum of four samples) over the previous 12 months. 'Free of infection' was defined as *S. aureus* negative over the previous 12 months (minimum of four samples). Samples were obtained as a part of routine clinical care.

2.2. Breath sampling

All patients followed the same procedure for breath collection. After rinsing their mouth with water, two breath samples were collected in 3 l Tedlar® bags (ProCare B.V., Groningen, the Netherlands) at a constant flow rate of 50 ml s^{-1}, using a commercial breath sampler (Loccioni, Angeli di Rosora, Italy) [19]. Each breath sample consisted of two mouth-exhalations through a bacterial filter (Air Safety Limited, Lancashire, UK) and a non-rebreathable T-piece (Vacumed, Ventura, USA) which were connected to a CO_2 sensor and a calibrated buffer pipe. From the total exhalation through the pipe, the first 150 ml of each exhalation was discarded in a separate small bag, and the remaining part collected in the sampling bag. The four exhalations were separated by 90 s, and at the same time, one sample of ambient air was collected. Equal volumes of breath and ambient air samples were stored within 6 h on glass tubes filled with Tenax TA® (Shimadzu, Kyoto, Japan), using a custom-made pump box at 3 l h^{-1}. This volume was determined by the lowest breath sample volume for each individual.

2.3. Breath analysis

Samples were analysed using thermal desorption (TD20) coupled to QP2010 Ultra GC-MS (Shimadzu, Kyoto, Japan) as described earlier [16]. Briefly, the tubes were dried in a custom-made setup at room temperature and thereafter heated in a TD unit for desorption of all trapped compounds. A cold trap captured the VOCs, and was subsequently heated to release the molecules through a heated transfer line (split 1:10) onto a CP-Sil 19 CF capillary column (25 m, 0.25 mm inner diameter, 1.2 μm film thickness, Agilent Technologies Netherlands BV, Amstelveen, the Netherlands). Helium was used as the carrier gas, at a flow-rate of 1.02 ml min^{-1}. The GC temperature program started isothermal at 40 °C for 4 min and was then ramped to 250 °C with a 5 °C min^{-1} heating rate. Finally, the temperature was kept isothermal at 250 °C for 14 min. After ionization of the molecules by electron impact, a quadrupole mass spectrometer detected nominal mass spectra with a scan rate of 20.000 amu s^{-1}. This resulted in data comprising of retention times (RTs) in a range 0–60 min (11 900 data points) and mass spectra in a range *m/z* 30–500, for each sample.

2.4. Compound identification

The compounds were putatively identified by comparing the experimental spectra with those of the National Institute for Standards and Technology (NIST) libraries NIST08 and NIST08s (80% minimal similarity). The assignment of the compounds most discriminating in the statistical analysis was performed with standards with

a purity of ⩾96% as described earlier [16]. The RTs and mass spectra of these pure standards were compared with those obtained in the experimental samples to confirm the correct identification of the compound (maximum difference in RT: 0.3% and a 90–95% MS match score, between chemical standards and biological samples). The compounds 1,4-pentadiene, hexanal, 2-methylnaphthalene and isopropyl myristate were purchased from Sigma Aldrich Chemie BV, Zwijndrecht, the Netherlands, ethanol was purchased from Merck KGaA, Darmstadt, Germany, 2-butanone was purchased from Merck Schuchardt, Hohenbrunn, Germany, acetone was purchased from VWR International BV, Amsterdam, the Netherlands, and 3-hydroxy-2-butanone was purchased from Thermo Fisher Scientific Inc., Karlsruhe, Germany. The standards were diluted 10 000 times in methanol (ethanol, acetone, 2-butanone, 3-hydroxy-2-butanone, hexanal, undecane, 2-methylnaphthalene, isopropyl myristate) or acetone (1,4-pentadiene) and 2 μl of this solution was injected onto the Tenax tube. Subsequently, the tubes were dried as described above and heated in the TD unit.

2.5. Multivariate analysis

2.5.1. Data preprocessing

The VOC data obtained was preprocessed using MetAlign [20] and subsequently using MSClust [21]. During preprocessing with MetAlign, the data underwent baseline correction, denoising, alignment and peak picking, i.e. molecular features extraction. During preprocessing with MSClust, molecular features from MetAlign were grouped together into clusters corresponding to the same compounds. Detailed settings of MetAlign and MSclust are available upon request. Oxygen and clusters attributed to Tedlar and Tenax (phenol, N,N-dimethylacetamide, and siloxane compounds) were removed from the data prior to analysis [22–24].

Samples were normalized by probabilistic quotient normalization (PQN) [25], a recommended method for normalization of chromatographic data [26]. Intensities of compounds were averaged for the duplicate breath samples. Correction for background concentrations of compounds in ambient air was performed by calculating alveolar gradients [27]: subtracting the corresponding ambient air sample from each average of two breath samples.

2.5.2. Discriminant analysis

Multivariate discriminant analysis was performed on the VOC profiles with s-PLS-DA [28] to discriminate *S. aureus* infected from non-infected CF patients. Partial Least Squares is an established method for dimension reduction. It does this by constructing latent variables (LVs), which are a relatively low number of linear combinations of all original variables (compounds). Sparse PLS integrates a variable reduction into this analysis by selecting a limited number of compounds to be represented in a LV [29].

Two parameters need to be optimized in s-PLS-DA, being the number of LVs in the PLS-DA model and the sparsity level that determines the maximum number of compounds that may contribute to each LV. These parameters were optimized simultaneously using double cross-validation [30]. Within this validation, an inner cross-validation, consisting of 5-fold cross-validation repeated 50 times, determined the optimal combination of number of LVs and sparsity level. The model performance was then determined by an outer cross-validation, also consisting of a 5-fold cross-validation repeated 50 times, as the median percentage of misclassified samples per class in all 50 repetitions. For cross-validation, samples were divided randomly, although stratified such that each test and training set contained at least one sample from each patient group. The overall optimal sparsity level was chosen to be the median of all optimal sparsity levels from the cross-validation models. The overall optimum number of LVs was taken to be the median number of LVs chosen for this optimal sparsity level. The statistical significance of the model performance was assessed by a permutation analysis with 1000 realizations using the optimal settings, in which the original class labels were randomly permuted. Compound importance was evaluated by the frequency of compound selection in different cross-validations and repetitions. If a compound was selected in at least 80% of all models, it was assumed to be relevant for the detection of *S. aureus* infection. For these compounds, the median intensities in *S. aureus* infected patients were compared to those in non-infected patients, to evaluate whether their concentrations were elevated or reduced. A Wilcoxon rank sum test was used to evaluate the difference between infected and non-infected patients for each of these compounds in a univariate manner. The threshold for significance was corrected for multiple testing using the Bonferroni correction [31] for an overall critical value of $\alpha = 0.05$.

With the exception of MetAlign and MSclust, all algorithms used in this paper were coded and executed in MATLAB version 8.3.0 (Mathworks, Natick MA). Sparse-PLS code was made available by Szymańska *et al* [32].

3. Results

We included 13 CF patients infected with *S. aureus* (10 with *S. aureus* on the day of breath sampling, and 3 with *S. aureus* in ⩾50% of their samples in the previous 12 months) and 5 CF patients free from *S. aureus* infection (table 1). One patient infected with *S. aureus* participated twice in the study, with 4 months in between sampling dates. It is very well possible for a patient to develop different infections in this period. Since this patient suffered from different infections during participation (see table 1, patient number 4), the breath profile is expected to differ significantly, and therefore these samples were treated as independent. This resulted in 19 measurement sessions with 2 breath

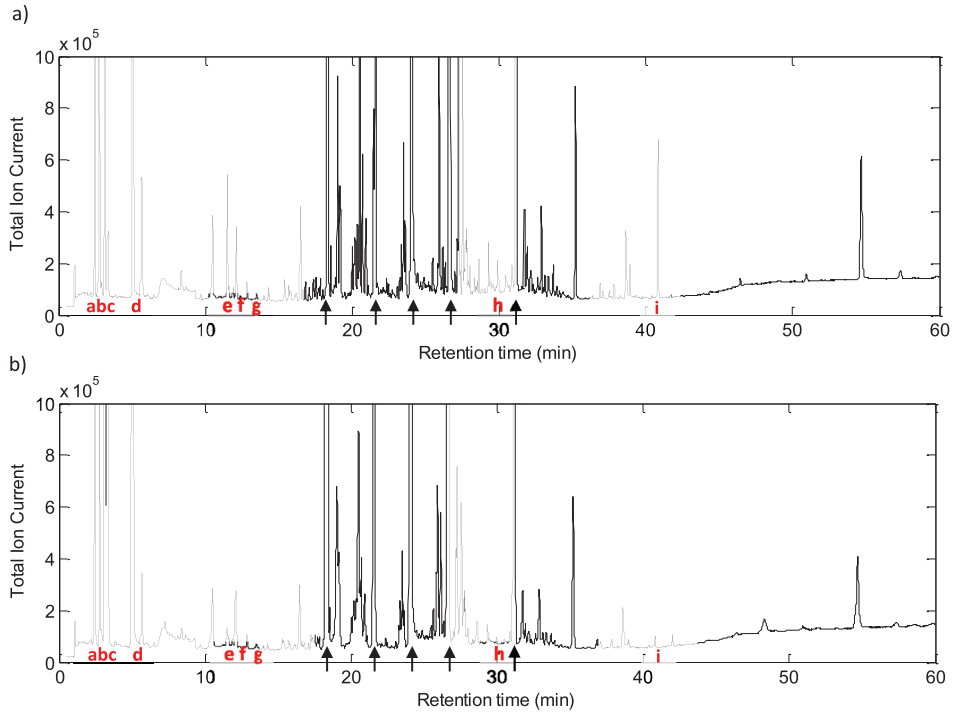

Figure 1. The average GC chromatogram of VOCs in exhaled breath of a *S. aureus* (a) non-infected and (b) infected CF patient. The horizontal axis represents the RT in minutes. The vertical axis (zoomed in to show also smaller peaks) shows the intensity (total ion current) of the compounds present in the sample. Arrows indicate peaks attributed to sample collection and adsorbent materials: phenol (24.6 min), N,N-dimethylacetamide (18.3 min), and siloxane compounds (21.6, 26.6, and 31.2 min). Letters correspond to compounds listed in table 2.

samples each. The median (interquartile range (IQR)) age in the *S. aureus* infected group was 13 (9–20) compared to 7 (7–13) in the group free from *S. aureus* infection.

Chromatograms were obtained for all breath and ambient air samples. The mean breath GC chromatograms (on a point-by-point basis) of *S. aureus* infected and non-infected CF patients before preprocessing are shown in figure 1. All peaks were within the linear range of the detector. Small differences can be observed between these chromatograms. The area between 15 and 35 min contains the most peaks. Five peaks, indicated with arrows, were attributed to compounds related to sample collection and adsorbent materials (Tedlar and Tenax, respectively) [22–24]. These peaks were removed prior to data analysis. Breath VOC profiles of *S. aureus* infected CF patients differ from those of non-infected patients. s-PLS-DA resulted in a median sensitivity of 100% and a specificity of 80.0%. Permutation testing showed that this prediction was significant ($p < 0.001$). The optimal parameter settings for s-PLS-DA were a sparsity level of 4 with 9 LVs.

The most important VOCs responsible for separation of infected and non-infected patients were identified as 1,4-pentadiene, ethanol, acetone, 2-butanone, 3-hydroxy-2-butanone, hexanal, undecane, 2-methylnaphthalene and isopropyl myristate, and are listed in table 2. The VOCs are correctly identified based on comparison with pure compound standards. As measure of importance, the frequency (expressed as a percentage) in which each compound was selected during cross-validation was determined. Ion counts of six compounds were elevated in concentration in *S. aureus* infected compared to non-infected patients, and reduced in concentration for three compounds. The significance threshold for the Wilcoxon rank sum test, after correcting for multiple testing, is $\alpha = 0.0056$

Table 2. Nine VOCs important for discrimination of *S. aureus* infected from non-infected CF patients, including their mass to charge ratio (*m/z*) and RT in minutes, using mean values determined in breath samples. The frequency in which each compound was selected during cross-validation is listed in the fifth column. The sixth column indicates whether the compound was elevated (↑) or reduced (↓) in *S. aureus* infected compared to non-infected CF patients. Column seven provides the *p*-values from the Wilcoxon rank sum test.

	RT	*m/z*	Name	%	Elevated (↑) or reduced (↓)	*p*-value
A	2.5	68	1,4-pentadiene	100	↑	0.010
B	2.8	46	Ethanol	84.0	↑	0.34
C	3.2	58	Acetone	85.6	↑	0.014
D	5.4	74	2-butanone	80.0	↑	0.50
E	11.5	88	3-hydroxy-2-butanone	100	↓	0.50
F	12.7	100	Hexanal	90.4	↓	0.69
G	13.6	156	Undecane	81.2	↑	0.0050
H	30.0	142	2-methyl-naphthalene	83.2	↑	0.75
I	40.8	270	Isopropyl myristate	99.2	↓	0.087

(=0.05/9). The compound undecane showed significant difference between non-infected and infected patients ($p = 0.0050$).

4. Discussion and conclusion

In this study, we investigated whether it was possible to discriminate *S. aureus* infected from non-infected CF patients based on VOCs in exhaled breath. Breath VOC profiles allowed successful detection of *S. aureus* infection, and this resulted in a shortlist of nine compounds important for the discrimination of the two patient groups. These nine compounds were correctly identified based on comparison with pure compound standards.

Previous studies have shown the potential of metabolomics of VOCs in exhaled breath of CF patients. Robroeks *et al* [33] identified CF patients with or without positive *P. aeruginosa* cultures 100% correctly using multiple discriminant analysis (MDA), by means of fourteen VOCs in exhaled breath. However, they did not report the molecular identity of those compounds. In addition, MDA builds a model using all measured compounds, many of which are possibly unrelated to *P. aeruginosa* infection. Including these compounds could weaken the discrimination power of the method and add noise to the model [34]. This also means that the trend of interest might be masked by irrelevant trends in the data (e.g. related to diet or physical activity) [35]. Sparse PLS-DA reduces these risks and facilitates biological interpretation by including only the most important compounds.

We identified the nine most discriminating compounds for *S. aureus* infection in CF patients. Almost all of these compounds have been reported previously either in bacterial headspace or in exhaled breath. The compounds 1,4-pentadiene, ethanol, acetone, 2-butanone, undecane and 2-methylnaphthalene were found to be elevated in *S. aureus* infected patients compared to non-infected. The compounds 3-hydroxy-2-butanone, hexanal and isopropyl myristate were reduced. One possible explanation for this would be uptake of these compounds by *S. aureus* bacteria. Additional research is required to investigate this.

Univariate analysis of the nine important compounds indicated undecane as significant, and, although undecane was also reported as a characteristic VOC in exhaled breath of lung cancer patients [36] and chronic obstructive pulmonary disease (COPD) patients [37], its use as a single biomarker deserves attention in further research. Analysis also revealed low p-values for 1,4-pentadiene, acetone and isopropyl myristate. The other compounds had relatively high p-values. Nonetheless, most of these compounds were frequently selected during cross-validation, indicating that they are important for a good discriminant model. This supports the use of multivariate analysis techniques, in which also the relationships between compounds are taken into account. In this way, also compounds that, on their own, do not show a difference between the groups can contribute greatly to the prediction of *S. aureus* infection [34].

The current 'golden standard' for detection of *S. aureus* in CF patients is a sputum culture. The reported sensitivity and specificity in adult patients is up to 100% and 63%, respectively [38]. However, young CF patients are not able to expectorate adequate samples and therefore upper airway swabs (including throat swabs) are used for diagnosis. The sensitivity of these swabs varies between 50–100%, and the specificity between 23–100% [38–42]. In our study, multivariate breath VOC profiles discriminated *S. aureus* infected from non-infected CF patients with a high sensitivity of 100% and a specificity of 80%, and therefore, breath analysis is an interesting alternative for current detection methods, especially in patients from whom a sputum sample is not available.

We are aware that five patients without *S. aureus* infection that were used as a reference is a relatively low number. However, this is a consequence of the strict criteria (free from *S. aureus* infection over the previous 12 months) we used for inclusion of patients in this group, whereas *S. aureus* infection is very common in CF patients. This means, unfortunately, that only a limited number of patients were free from *S. aureus* infection, and could be included in this study. Secondly, CF patients very often have respiratory infections. Therefore, these patients might be (co-)infected with other pathogens, and we previously showed in an *in vitro* study [16] that VOC profiles in co-microbial environment differ from those released by pathogens in monoculture. For this reason, the VOCs in exhaled breath of patients with mixed respiratory infections may be different as well.

In general, breath profiles are influenced by many (confounding) factors which are difficult to control. For example, diet, medication, and the time of day the sample is taken. We are aware that the small number of patients included in this study enlarges the influence of these factors on the difference between patient groups detected by s-PLS-DA. Also exogenous compounds, such as those present in ambient air, can be confounding factors. However, we subtracted the compound concentrations in ambient air from those in exhaled breath to correct for this, as described by Philips *et al* [27]. It is true that this procedure does not take into account the complexity of gas exchange processes in the lung such as the alveolar concentration gradients of the substances [43]. However, the best alternative would be to let participants breathe pure air before they provide a breath sample, which is time consuming and therefore not suitable for clinical routine measurements [44]. To reduce the influence of breath-to-breath variability we averaged duplicate measurements. Furthermore, a

strict cross-validation scheme kept the effect of individual variability to a minimum. This study aimed to determine whether we can discriminate between *S. aureus* infected from non-infected CF patients. Yet, for the identification of specific biomarkers for *S. aureus* infection larger cohort studies are needed.

We conclude from this proof of principle study that GC-MS breath VOC profiles are promising for detection of *S. aureus* infections in CF patients and that s-PLS-DA is a useful tool in such studies. With s-PLS-DA we were able to discriminate *S. aureus* infected from non-infected patients with 100% sensitivity and 80% specificity. We obtained a shortlist of nine important compounds for the detection of *S. aureus* that can be used as a focus point in further biomarker identification research. In addition, this study may inspire and lead to future studies into the non-invasive detection of others airway infections in CF patients.

Acknowledgments

The authors would like to thank all the study participants and their parents. Furthermore, we would like to acknowledge all (research) nurses for their help in organizing the collection of exhaled breath samples. This work was funded by the Longfonds (The Netherlands, Project No. 3.3.11.002).

Conflicts of interest

No author reported any conflict of interest.

References

[1] O'Sullivan B P and Freedman S D 2009 Cystic fibrosis *Lancet* **373** 1891–904
[2] Cystic Fibrosis Foundation 2011 *Annual Data Report* (Bethesda, MD: Cystic Fibrosis Foundation)
[3] Al-Saleh S, Dell S D, Grasemann H, Yau Y C, Waters V, Martin S and Ratjen F 2010 Sputum induction in routine clinical care of children with cystic fibrosis *J. Pediatr.* **157** 1006–11.e1
[4] Armstrong D S, Grimwood K, Carlin J B, Carzino R, Gutierrez J P, Hull J, Olinsky A, Phelan E M, Robertson C F and Phelan P D 1997 Lower airway inflammation in infants and young children with cystic fibrosis *Am. J. Respir. Crit. Care Med.* **156** 1197–204
[5] De Boeck K, Alifier M and Vandeputte S 2000 Sputum induction in young cystic fibrosis patients *Eur. Respir. J.* **16** 91–4
[6] Graffelman A W, Willemssen F E, Zonderland H M, Neven A K, Kroes A C and van den Broek P J 2008 Limited value of chest radiography in predicting aetiology of lower respiratory tract infection in general practice *Br. J. Gen. Pract.* **58** 93–7
[7] Bos L D, Sterk P J and Schultz M J 2013 Volatile metabolites of pathogens: a systematic review *PLoS Pathog.* **9** e1003311
[8] Filipiak W, Sponring A, Baur M M, Filipiak A, Ager C, Wiesenhofer H, Nagl M, Troppmair J and Amann A 2012 Molecular analysis of volatile metabolites released specifically by *Staphylococcus aureus* and *Pseudomonas aeruginosa BMC Microbiol.* **12** 113
[9] Chippendale T W, Gilchrist F J, Španěl P, Alcock A, Lenney W and Smith D 2014 Quantification by SIFT-MS of volatile compounds emitted by *in vitro* cultures of *S. aureus*, *S. pneumoniae* and *H. influenzae* isolated from patients with respiratory diseases *Anal. Methods* **6** 2460–72
[10] Zhu J, Bean H D, Kuo Y M and Hill J E 2010 Fast detection of volatile organic compounds from bacterial cultures by secondary electrospray ionization-mass spectrometry *J. Clin. Microbiol.* **48** 4426–31
[11] Boots A, Smolinska A, van Berkel J, Fijten R, Stobberingh E, Boumans M, Moonen E, Wouters E, Dallinga J and Van Schooten F 2014 Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography-mass spectrometry *J. Breath Res.* **8** 027106
[12] Tait E, Perry J D, Stanforth S P and Dean J R 2014 Identification of volatile organic compounds produced by bacteria using HS-SPME-GC-MS *J. Chromatogr. Sci.* **52** 363–73
[13] Chippendale T W, Gilchrist F J, Španěl P, Alcock A, Lenney W and Smith D 2014 Quantification by SIFT-MS of volatile compounds emitted by *Aspergillus fumigatus* cultures and in co-culture with *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Streptococcus pneumoniae Anal. Methods* **6** 8154–64
[14] Filipiak W, Sponring A, Baur M M, Ager C, Filipiak A, Wiesenhofer H, Nagl M, Troppmair J and Amann A 2012 Characterization of volatile metabolites taken up by or released from *Streptococcus pneumoniae* and *Haemophilus influenzae* by using GC-MS *Microbiology* **158** 3044–53
[15] Farag M A, Ryu C-M, Sumner L W and Paré P W 2006 GC-MS SPME profiling of rhizobacterial volatiles reveals prospective inducers of growth promotion and induced systemic resistance in plants *Phytochemistry* **67** 2262–8
[16] Neerincx A, Geurts B, Habets M, Booij J, van Loon J, Jansen J, Buydens L, van Ingen J, Mouton J and Harren F 2016 Identification of *Pseudomonas aeruginosa* and *Aspergillus fumigatus* mono-and co-cultures based on volatile biomarker combinations *J. Breath Res.* **10** 016002
[17] Bean H D, Jiménez-Díaz J, Zhu J and Hill J E 2015 Breathprints of model murine bacterial lung infections are linked with immune response *Eur. Respir. J.* **45** 181–90
[18] Filipiak W, Beer R, Sponring A, Filipiak A, Ager C, Schiefecker A, Lanthaler S, Helbok R, Nagl M and Troppmair J 2015 Breath analysis for *in vivo* detection of pathogens related to ventilator-associated pneumonia in intensive care patients: a prospective pilot study *J. Breath Res.* **9** 016004
[19] Thornadtsson A, Neerincx A, Högman M, Hugen C, Sintnicolaas C, Harren F, Merkus P and Cristescu S 2015 Extended nitric oxide analysis may improve personalized anti-inflammatory treatment in asthmatic children with intermediate FENO50 *J. Breath Res.* **9** 047114
[20] Lommen A 2009 MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing *Anal. Chem.* **81** 3079–86
[21] Tikunov Y, Laptenok S, Hall R, Bovy A and De Vos R 2012 MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data *Metabolomics* **8** 714–8
[22] Reynolds J C, Jimoh M A, Guallar-Hoyas C, Creaser C S, Siddiqui S and Thomas C L P 2014 Analysis of human breath samples using a modified thermal desorption: gas chromatography electrospray ionization interface *J. Breath Res.* **8** 037105
[23] Steeghs M M, Cristescu S M and Harren F J 2006 The suitability of Tedlar bags for breath sampling in medical diagnostic research *Physiol. Meas.* **28** 73
[24] Uhde E 1999 Application of solid sorbents for the sampling of volatile organic compounds in indoor air *Organic Indoor Air Pollutants: Occurrence, Measurement, Evaluation* 2nd edn, ed T Salthammer and E Uhde (Weinheim: Wiley) pp 1–14
[25] Dieterle F, Ross A, Schlotterbeck G and Senn H 2006 Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics *Anal. Chem.* **78** 4281–90
[26] Filzmoser P and Walczak B 2014 What can go wrong at the data normalization step for identification of biomarkers? *J. Chromatogr.* A **1362** 194–205

[27] Phillips M 1997 Method for the collection and assay of volatile organic compounds in breath *Anal. Biochem.* **247** 272–8
[28] Lê Cao K-A, Boitard S and Besse P 2011 Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems *BMC Bioinformatics* **12** 1
[29] Chun H and Keleş S 2010 Sparse partial least squares regression for simultaneous dimension reduction and variable selection *J. R. Stat. Soc.* B **72** 3–25
[30] Szymańska E, Saccenti E, Smilde A K and Westerhuis J A 2012 Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies *Metabolomics* **8** 3–16
[31] Bland J M and Altman D G 1995 Multiple significance tests: the Bonferroni method *BMJ* **310** 170
[32] Szymańska E, Tinnevelt G H, Brodrick E, Williams M, Davies A N, van Manen H-J and Buydens L M 2016 Increasing conclusiveness of clinical breath analysis by improved baseline correction of multi capillary column-ion mobility spectrometry (MCC-IMS) data *J. Pharm. Biomed. Anal.* **127** 170–5
[33] Robroeks C M M H T *et al* 2010 Metabolomics of volatile organic compounds in cystic fibrosis patients and controls *Pediatr. Res.* **68** 75–80
[34] Szymańska E, Gerretzen J, Engel J, Geurts B, Blanchet L and Buydens L M 2015 Chemometrics and qualitative analysis have a vibrant relationship TrAC *Trends Anal. Chem.* **69** 34–51
[35] Brereton R G 2006 Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data TrAC *Trends Anal. Chem.* **25** 1103–11
[36] Phillips M, Gleeson K, Hughes J M B, Greenberg J, Cataneo R N, Baker L and McVay W P 1999 Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study *Lancet* **353** 1930–3
[37] Van Berkel J, Dallinga J, Möller G, Godschalk R, Moonen E, Wouters E and Van Schooten F 2010 A profile of volatile organic compounds in breath discriminates COPD patients from controls *Respir. Med.* **104** 557–63
[38] Seidler D, Griffin M, Nymon A, Koeppen K and Ashare A 2016 Throat swabs and sputum culture as predictors of *P. aeruginosa* or *S. aureus* lung colonization in adult cystic fibrosis patients *PLoS One* **11** e0164232
[39] Armstrong D S, Grimwood K, Carlin J B, Carzino R, Olinsky A and Phenlan P D 1996 Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis *Pediatr. Pulmonol.* **21** 267–75
[40] Kabra S, Alok A, Kapil A, Aggarwal G, Kabra M, Lodha R, Pandey R, Sridevi K and Mathews J 2004 Can throat swab after physiotherapy replace sputum for identification of microbial pathogens in children with cystic fibrosis? *Indian J. Pediatr.* **71** 21–3
[41] Ramsey B W, Wentz K R, Smith A L, Richardson M, Williams-Warren J, Hedges D L, Gibson R, Redding G J, Lent K and Harris K 1991 Predictive value of oropharyngeal cultures for identifying lower airway bacteria in cystic fibrosis patients *Am. Rev. Respir. Dis.* **144** 331–7
[42] Rosenfeld M, Emerson J, Accurso F, Armstrong D, Castile R, Grimwood K, Hiatt P, McCoy K, McNamara S and Ramsey B 1999 Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis *Pediatr. Pulmonol.* **28** 321–8
[43] Wagner P D, Saltzman H A and West J B 1974 Measurement of continuous distributions of ventilation-perfusion ratios: theory *J. Appl. Physiol.* **36** 588–99
[44] Risby T H and Sehnert S S 1999 Clinical application of breath biomarkers of oxidative stress status *Free Radic. Biol. Med.* **27** 1182–92