Journal of Breath Research

PAPER

Preliminary method for profiling volatile organic compounds in breath that correlate with pulmonary function and other clinical traits of subjects diagnosed with cystic fibrosis: a pilot study

M Woollam[1,2], A P Siegel[1,2], P Grocki[1,2], J L Saunders[3], D B Sanders[3], M Agarwal[1,2,4,*] and M D Davis[3,5,*]

[1] Integrated Nanosystems Development Institute Indiana University-Purdue University of Indianapolis, Indianapolis, IN, United States of America
[2] Department of Chemistry and Chemical Biology Indiana University-Purdue University of Indianapolis, Indianapolis, IN, United States of America
[3] Division of Pulmonology, Allergy, and Sleep Medicine Riley Hospital for Children at Indiana University School of Medicine, Indianapolis, IN, United States of America
[4] Department of Mechanical and Energy Engineering Indiana University-Purdue University of Indianapolis, Indianapolis, IN, United States of America
[5] Wells Center for Pediatric Research Riley Hospital for Children at Indiana University School of Medicine, Indianapolis, IN, United States of America
[*] Authors to whom any correspondence should be addressed.

E-mail: MDD1@iu.edu **and** agarwal@iupui.edu

Keywords: exhaled breath, exhaled VOCs, cystic fibrosis

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
19 November 2021

REVISED
27 January 2022

ACCEPTED FOR PUBLICATION
4 February 2022

PUBLISHED
22 February 2022

Abstract

Cystic fibrosis (CF) is characterized by chronic respiratory infections which progressively decrease lung function over time. Affected individuals experience episodes of intensified respiratory symptoms called pulmonary exacerbations (PEx), which in turn accelerate pulmonary function decline and decrease survival rate. An overarching challenge is that there is no standard classification for PEx, which results in treatments that are heterogeneous. Improving PEx classification and management is a significant research priority for people with CF. Previous studies have shown volatile organic compounds (VOCs) in exhaled breath can be used as biomarkers because they are products of metabolic pathways dysregulated by different diseases. To provide insights on PEx classification and other CF clinical factors, exhaled breath samples were collected from 18 subjects with CF, with some experiencing PEx and others serving as a baseline. Exhaled breath was collected in Tedlar bags during tidal breathing and cryotransferred to headspace vials for VOC analysis by solid phase microextraction coupled to gas chromatography–mass spectrometry. Statistical significance testing between quantitative and categorical clinical variables displayed percent-predicted forced expiratory volume in one second (FEV1pp) was decreased in subjects experiencing PEx. VOCs correlating with other clinical variables (body mass index, age, use of highly effective modulator treatment (HEMT), and the need for inhaled tobramycin) were also explored. Two volatile aldehydes (octanal and nonanal) were upregulated in patients not taking the HEMT. VOCs correlating to potential confounding variables were removed and then analyzed by regression for significant correlations with FEV1pp measurements. Interestingly, the VOC with the highest correlation with FEV1pp (3,7-dimethyldecane) also gave the lowest *p*-value when comparing subjects at baseline and during PEx. Other VOCs that were differentially expressed due to PEx that were identified in this study include durene, 2,4,4-trimethyl-1,3-pentanediol 1-isobutyrate and 5-methyltridecane. Receiver operator characteristic curves were developed and showed 3,7-dimethyldecane had higher

ability to classify PEx (area under the curve (AUC) = 0.91) relative to FEV1pp values at collection (AUC = 0.83). However, normalized ΔFEV1pp values had the highest capability to distinguish PEx (AUC = 0.93). These results show that VOCs in exhaled breath may be a rich source of biomarkers for various clinical traits of CF, including PEx, that should be explored in larger sample cohorts and validation studies.

1. Introduction

1.1. Cystic fibrosis—background, treatment, and priorities

Cystic fibrosis (CF) is a multisystem autosomal recessive disease that, from early life, leads to deficits in growth, nutrition and progressive lung disease [1, 2]. Lung disease is the leading cause of morbidity and mortality for people with CF and is characterized by impaired mucociliary clearance, persistent microbial infection, and an exaggerated inflammatory response [3]. In addition, pulmonary exacerbations (PEx) are a common complication in this population. PEx are defined as episodes of acute or subacute clinical worsening characterized by increased cough, increased sputum production, loss of lung function, and weight loss among other signs and symptoms [4]. PEx treated with intravenous (IV) antibiotics are associated with increased mortality, declines in percent predicted forced expiratory volume in one second (FEV1pp), poorer quality of life, increased healthcare costs, and reduced survival [5–10]. The frequency of courses of oral and inhaled antibiotics is relatively unknown, but PEx treated with these therapies are even more common than PEx treated with IV antibiotics [11]. Despite their frequent occurrences, our understanding of the causes of PEx is limited. Various etiologies have been investigated, including acute events (e.g. viral infections [12], clonal shifts of colonizing bacteria [13], acute environmental exposures [14]), progression of underlying disease associated with medical non-adherence [15], increasing infection burden [16], and others [17, 18]. Furthermore, the recognition and treatment of PEx is highly variable.

The introduction of highly effective modulator treatment (HEMT) for people with CF has greatly reduced the frequency of PEx but may not have eliminated the associated poor outcomes. Deciding whether to use antibiotics to treat new respiratory symptoms of people with CF taking HEMT may become more difficult, especially since most PEx are treated over the phone without measuring change in FEV1pp. FEV1pp can be measured at home using home spirometry devices, but these tend to produce more variable results and further complicate treatment decisions. Having a simpler point-of-care measure that can help identify impending PEx that would benefit from treatment would be greatly beneficial. Such a measure could both reduce the overuse of antibiotics and potential for increasing resistance, and the under-recognition of PEx to enable early treatment and recovery of lung function.

1.2. Exhaled volatile organic compounds and CF

Because of its non-invasive and virtually unlimited nature, interest in exhaled breath as a matrix for biomarkers has greatly increased over the past two decades. Contained in exhaled breath are hundreds of volatile organic compounds (VOCs) that are in vapor-phase at body temperature (37°). VOCs expressed in exhaled breath (and other biological matrices) have been investigated for many different diseases [19–23]. The biochemical rationale for utilizing VOCs in exhaled breath as biomarkers is that they are biological products of metabolic pathways that are uniquely altered by the initiation and progression of diseases [24, 25]. Additionally, VOCs expressed in breath could also be produced microbially by the presence of proliferating bacteria [26–28]. Of note, canines have demonstrated they can detect an array of diseases just by smelling VOCs emanating from breath and other biological samples [29, 30]. There are many analytical methods that can be used to quantitatively profile VOCs in breath. The current gold standard technique used for VOC biomarker discovery is gas chromatography–mass spectrometry (GC-MS) and is regarded as such because it can chromatographically separate the VOCs, identify their molecular structure by their individual ionization patterns (mass spectra), and quantitate their concentrations in biological samples [31]. Solid phase microextraction (SPME) is often coupled to GC-MS for VOC biomarker discovery [32–35], as SPME can preconcentrate VOCs prior to GC-MS analysis and increase the sensitivity for VOC detection in breath samples.

Exhaled VOCs have been shown to differentiate people with CF from healthy control subjects and have been correlated with the presence of specific bacterial infections and PEx [26–28, 36, 37]. Nonetheless, even though VOCs have been explored as a potential source of biomarkers, comprehensive VOC analyses relating to an array of key CF clinical variables have not been investigated. These unexplored variables include confounders such as growth and nutrition (typically using body mass index (BMI) and age), as well as other clinical factors such as the chronic use of inhaled antibiotics, the effects of HEMT and most importantly, baseline values of FEV1pp. We

hypothesize that exhaled VOCs will correlate with lung function and other clinical traits of CF. We further hypothesize that exploring these correlations will help identify VOCs that are sensitive or specific for PEx. Thorough investigation of exhaled VOCs in CF could allow for the acceleration of breath analysis to reach the clinic and potentially provide beneficial outcomes to patients. Finally, utilizing exhaled VOCs to classify PEx in subjects with CF could be superior to, or complement, the current standard-of-care diagnostic techniques.

2. Materials and methods

2.1. Materials and instrumentation

Parafilm, reagent alcohol and ViroMax filters were purchased from Fisher Scientific USA (Florence, KY). A mass flow controller (flowmeter) was purchased from Alicat (Tucson, AZ). About 20 ml headspace vials with screw cap lids, deactivated glass wool, 3 l Tedlar gas sampling bags, and 1 cm polydimethylsiloxane/carboxen/divinylbenzene (PDMS/CAR/DVB) SPME fibers were manufactured by Restek (Bellefonte, PA). Stainless steel needles (Med-Vet International, Mettawa, IL) were used for cryotranferring VOCs to headspace vials and were cleaned between uses. VOCs sampled and concentrated by the SPME fiber were analyzed using an Agilent (Santa Clara, CA) 9000 Intuvo GC system coupled to an Agilent 5977B single quadrupole MS system. The column utilized was an Agilent Ultra Inert HP-5 ms, 5% phenyl methyl siloxane GC column of 30 m length, 250 μm internal diameter, and 0.25 μm film thickness.

2.2. Study design/subject info/clinical data

Subjects between the ages of 8 and 18 years with a diagnosis of CF attending clinic visits or admitted to Riley Hospital for Children were eligible to enroll in this study. A convenience sample of subjects at their baseline lung health was obtained as well as from subjects experiencing a PEx. PEx was defined as the treating clinicians' choice to treat with antibiotics for new respiratory symptoms and/or a decline of FEV1pp >10% predicted from each individual subject's baseline. Baseline FEV1pp was defined as the average of the best two FEV1pp measurements in the year prior to the clinic visit. Subjects were eligible to enroll more than once. Demographic information including age, sex, and medical history (genotype, CF comorbidities, medications) were recorded.

2.3. Ethical statement

The Indiana University Institutional Review Board approved this study (IRB # 12005 and 1910580775) and it was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. Informed consent and assent, when appropriate, were obtained from all patients.

2.4. Breath collection

Breath samples were collected into Tedlar bags using a modified procedure which has been previously published [35]. Subjects breathed tidally through a ViroMax viral filter (viral filtration efficiency greater than 99.99%) coupled to a 3 l Tedlar bag through a small inlet, until the bag was 80% full. Next, the researchers placed the bags in secondary polypropylene bags which were sanitized with reagent alcohol (70%) prior to transportation back to the instrumentation facility for analysis. Previous implementation of this method (prior to the COVID-19 pandemic) did not include use of a viral filter or addition of a secondary container that could be sanitized with reagent alcohol. Therefore, the modified protocol was compared with the previously implemented method quantitatively, to ensure the safety modifications did not impede the detection of VOCs in breath samples. There was no significant difference in the total number of VOCs detected or the total integrated signal between the previous method and the modified method (relative standard deviation of the total signal was equal to 9.9%, supplementary figure S1 available online at stacks.iop.org/JBR/16/027103/mmedia).

2.5. Sample processing and preservation

VOCs were cryotransferred to a headspace vial containing glass wool [35]. Briefly, headspace vial cap septa were pierced by two needles, one long and one short, to enable VOCs to flow through the vials. The method entailed cooling the vial to approximately −45 °C using dry ice and then using a vacuum interfaced with a flowmeter (attached to the shorter needle) to transfer the air slowly from the Tedlar bags (attached to the longer needle), depositing the VOCs onto the extremely cold glass wool and the inner walls of the headspace vial. The headspace vials were wrapped with parafilm and stored at −80 °C, and subsequently analyzed by SPME GC-MS QTOF. An illustration of sample collection, the cryotransfer process and SPME GC-MS QTOF is shown in figure 1.

2.6. SPME GC-MS analysis

One PDMS/CAR/DVB SPME fiber was utilized for all runs. Prior to the daily runs, the PDMS/CAR/DVB fiber was preconditioned at 250 °C for 10 min. Vials were warmed to room temperature and incubated in a 60 °C water bath for 45 min in the presence of the SPME fiber, which was manually injected into the GC-MS system. Daily reference calibration standards were utilized to track instrumental variability and normalize sample output. The oven temperature program utilized an initial temperature of 40 °C held for 2 min, followed by a ramp of 8 °C min^{-1}–100 °C, followed by 15 °C min^{-1}–120 °C, followed by

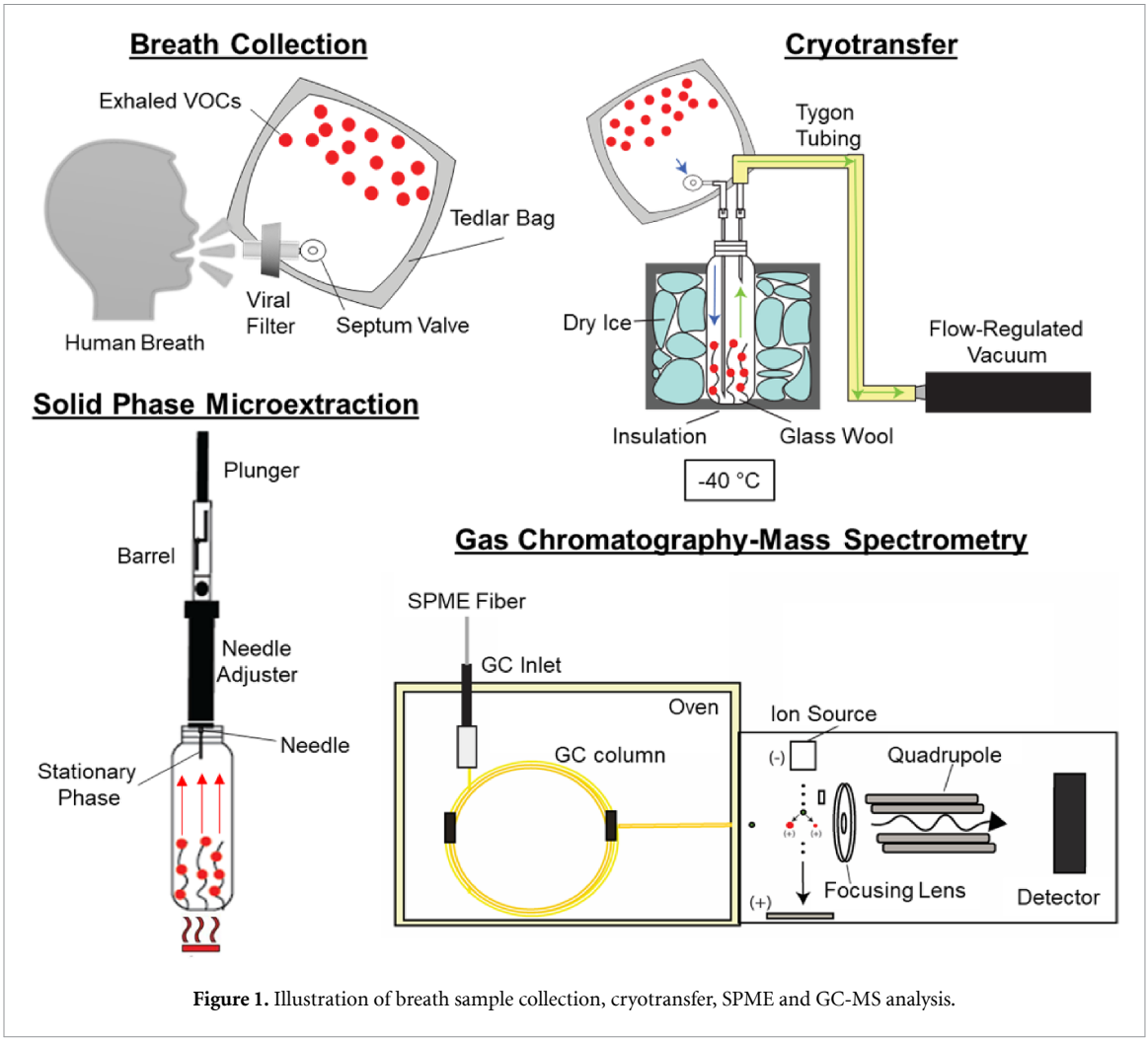

Figure 1. Illustration of breath sample collection, cryotransfer, SPME and GC-MS analysis.

8 °C min^{-1}–180 °C, 15 °C min^{-1}–200 °C, and 8 °C min^{-1}–260 °C. The MS transfer line temperature was 250 °C during the entirety of the chromatographic run.

2.7. Data processing and chemometric analysis

All sample chromatograms were spectrally aligned using Mass Hunter Profinder (Agilent proprietary software) based on similarities in retention time and mass spectral fragmentation patterns to identify conserved VOCs in samples. The spectrally aligned matrix was filtered for the removal of VOCs with large retention time variation and VOCs not detected in at least 50% of the samples in one class for each statistical comparison. This was performed to remove VOCs that were not consistently detected in samples, as these molecules will not be reliable biomarkers for CF clinical traits. Additionally, silanes and siloxanes were disqualified from the data matrix. Next, VOC signals were normalized relative to the daily reference standards to account for instrumental variation. After normalization, VOCs displaying large instrumental variation were disqualified from analysis. Next, normalized VOC signals were analyzed for significant correlations with variables including age, BMI, and the use of HEMT (Trikafta (elexacaftor/tezacaftor/ivacaftor), Orkambi (lumacaftor/ivacaftor), or Symdeko (tezacaftor/ivacaftor)) or antibiotics (specifically inhaled tobramycin). VOCs were also probed for significant correlations with FEV1pp (prognostic factor for PEx) and PEx. Significance testing was undertaken using the Mann-Whitney U-test (categorical clinical variables). On the other hand, analysis of variance (ANOVA) testing was utilized to discover significant VOC correlations (quantitative clinical variables).

3. Results

3.1. Descriptive results

A total of 18 breath samples were collected at Riley Hospital for Children at Indiana University School of Medicine from subjects (aged 8–18) diagnosed with CF. A detailed description of subjects is shown in table 1. Clinical traits include PEx, FEV1pp at collection, change in FEV1pp relative to baseline levels (ΔFEV1pp), the need for inhaled tobramycin at the time of collection, use of the HEMT, age and BMI. There were seven PEx and 11 baseline subjects in this

Table 1. List of CF patients consented in this study and associated clinical traits of interest.

	Age	Baseline FEV1pp	Collection FEV1pp	ΔFEV1pp	BMI %ile	Inhaled tobramycin	HEMT	Clinical status
Subject 1	13	88.5	98	9.5	59	Yes	No	Baseline
Subject 2	17	90.5	74	−16.5	77	No	No	PEx
Subject 3	9	122	126	4	99	No	Yes	Baseline
Subject 4	10	110	92	−18	99	Yes	Yes	Baseline
Subject 5	17	85	90	5	64	No	No	Baseline
Subject 6	14	70	51	−19	55	Yes	Yes	PEx
Subject 7	14	96	102	6	49	No	No	Baseline
Subject 8	16	107.5	77	−30.5	70	Yes	Yes	PEx
Subject 9	14	110	103	−7	10	Yes	No	PEx
Subject 10	12	83	83	0	77	No	Yes	Baseline
Subject 11	10	85	89	4	34	No	Yes	Baseline
Subject 12	10	99.5	89	−10.5	65	No	No	Baseline
Subject 13	14	99.5	96	−3.5	59	No	Yes	Baseline
Subject 14	14	103	101	−2	90	No	Yes	Baseline
Subject 15	18	87.5	54	−33.5	67	No	No	PEx
Subject 16	12	82.5	75	−7.5	33	Yes	Yes	Baseline
Subject 17	14	81	66	−15	46	Yes	Yes	PEx
Subject 18	8	106.5	86	−20.5	74	Yes	Yes	PEx

sample cohort. Among the subjects, 11/18 subjects were taking HEMT, and 8/18 were chronically administered tobramycin via inhalation. Most samples were collected in the outpatient clinic as only two samples were collected from the inpatient ward (subject 8 and subject 15). Prior to profiling any VOCs, the team analyzed BMI, age, and FEV1pp for significant differences between the categorical variables (PEx, HEMT, inhaled tobramycin). Box and whisker plots for each comparison can be observed in figure 2. The only statistically significant differences between PEx and baseline were in FEV1pp (mean = 86.2, standard deviation = 18.3, p-value = 0.02) and ΔFEV1pp (mean = −8.6, standard deviation = 12.7, p-value = 0.0007).

3.2. Removal of VOCs that correlate with age and BMI

There were 125 qualified VOCs after spectrally aligning the chromatograms and filtering the data (removing silanes/siloxanes, non-endogenous VOCs, etc.). The team decided to only analyze VOCs from specific functional groups, namely carbonyls, hydrocarbons and volatile terpenes/terpenoids. The rationale is that these compounds are potentially endogenous and have been previously reported biomarkers of PEx [37] and other diseases [32–34, 38]. This left a total of 74 VOCs for analysis. VOCs were normalized and initially analyzed for correlations with age and BMI. Two VOCs were found to significantly correlate with BMI, one showing a positive correlation (octyl acetate, $r = 0.62$, p-value = 0.005) and the other displaying negative correlation (3-methylundecane, $r = 0.48$, $p = 0.04$). Additionally, two volatile compounds significantly correlated with CF subject age. Again, one of these VOCs showed a negative correlation (isocaryophyllene, $r = 0.50$, $p = 0.04$) and the other demonstrated a positive correlation (octanal, $r = 0.59$, p-value = 0.009).

3.3. Analyzing the effects of the use of chronic inhaled tobramycin and HEMT on exhaled VOCs

VOCs were also analyzed for differences associated with the use of inhaled tobramycin and HEMT. Only one VOC was found to be dysregulated in subjects with CF who were chronically administered inhaled tobramycin. This VOC was identified as 5-(2-methylpropyl)-nonane (mean normalized signal = 0.0008, standard deviation = 0.0009 and p-value = 0.02) and was enriched in the breath of subjects with CF who were not chronically taking inhaled tobramycin. VOCs were analyzed for differences between subjects on HEMT and those who were not. Regardless of univariate statistical significance, most VOCs were upregulated in subjects not taking HEMT (see volcano plot in figure 3(a)). Interestingly, four out of the qualified VOCs had $p < 0.05$ and 12 VOCs displayed $p < 0.10$ (Mann–Whitney U-test). The 12 VOCs can be observed in the heatmap in figure 3(b). The columns represent CF breath samples, and the rows depict VOCs. The first seven VOCs demonstrate high interclass variation with low intraclass variation. The last four VOCs in the heatmap demonstrate low variation within subjects taking HEMT, but relatively high variation among subjects not taking HEMT. Three VOCs with relatively high ability to discriminate subjects based on the use of HEMT are illustrated in figures 3(c)–(e). These VOCs include a saturated hydrocarbon (2,2,8-trimethyldecane, mean normalized signal = 0.0002, standard deviation = 0.0004, p-value = 0.02) and two volatile aldehydes (octanal (mean normalized signal = 0.0007, standard deviation = 0.0005, p-value = 0.04) and nonanal (mean normalized signal = 0.010, standard deviation = 0.004, p-value = 0.02)). All three of

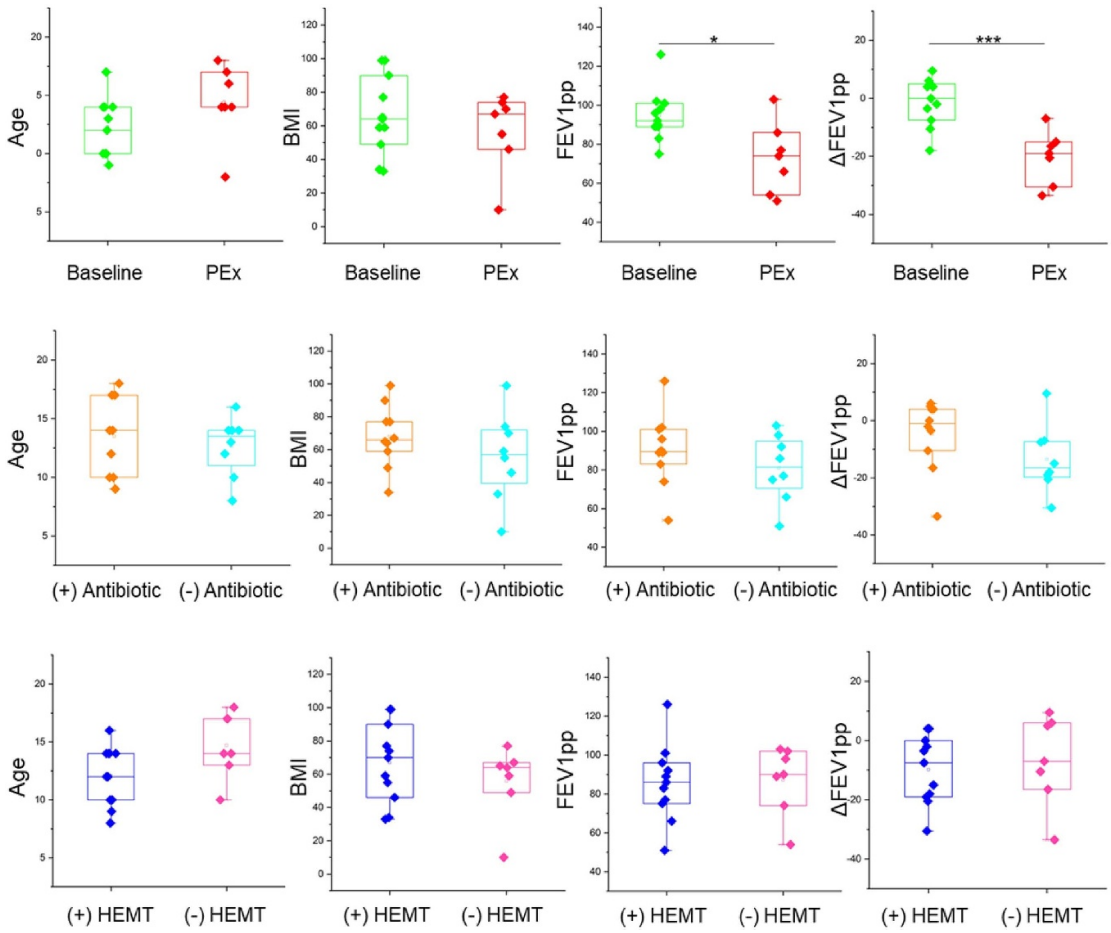

Figure 2. Box and whisker plots of the quantitative variables (FEV1pp, BMI, age) among the different categorical variables (PEx, inhaled tobramycin, HEMT). Significant differences in FEV1pp were identified during PEx.

these VOCs were upregulated in subjects not taking HEMT. The other VOC with $p < 0.05$ was identified as 2-methylpenyl ester cyclohexane carboxylic acid (mean normalized signal = 0.0001, standard deviation = 0.0002, p-value = 0.03) and was also upregulated. VOCs that were correlated with age, BMI and the use of HEMT and need for inhaled tobramycin were discarded from further analysis, where VOCs were analyzed for correlations with pulmonary function. This is because none of these clinical factors had significant effects on FEV1pp or PEx in this study (see table 1 and figure 2).

3.4. Correlating VOCs with FEV1pp and PEx

After removing VOCs correlated with age, BMI, and use of HEMT or inhaled tobramycin, 66 VOCs were analyzed using linear regression to observe any correlations with FEV1pp at the time of collection. Four of the VOCs were identified to have significant correlations. These VOCs were identified as 3,7-dimethyldecane ($r = 0.64$, p-value = 0.004), terpinen-4-ol ($r = 0.56$, p-value = 0.01), 4-methyl-octane ($r = 0.49$, p-value = 0.04) and 4,6-dimethyldodecane ($r = 0.48$, p-value = 0.04). The VOC with the strongest correlation with FEV1pp (3,7-dimethyldecane) can be observed in figure 4(a). ΔFEV1pp was also analyzed for correlations with VOCs, and the VOC with the highest correlation again was 3,7-dimethyldecane. This VOC is also shown in the bottom panel of figure 4(a) and demonstrated a strong correlation ($r = 0.66$, $p = 0.003$). 4-methyl-octane also correlated with FEV1pp and ΔFEV1pp ($r = 0.48$, $p = 0.04$). The VOCs correlating with FEV1pp were not removed from further analysis, as FEV1pp is used as a prognostic factor for PEx and FEV1pp values were significantly lower in subjects experiencing PEx in this sample cohort (see box/whisker plots in figure 2).

Next, VOCs were probed for significant differences in subjects experiencing PEx using the Mann–Whitney U-test. Four of the qualified VOCs had $p < 0.05$ and the VOC with the lowest p-value was 3,7-dimethyldecane (box/whisker plot for this VOC can be seen in figure 4(b)) which was the VOC with the strongest correlation to FEV1pp and ΔFEV1pp (figure 4(a)). This was the only VOC that was significantly correlated with FEV1pp and differentially expressed in subjects experiencing PEx. 3,7-dimethyldecane was also significantly downregulated during PEx alongside several other VOCs (figure 4(c)). The four VOCs which were dysregulated during PEx were

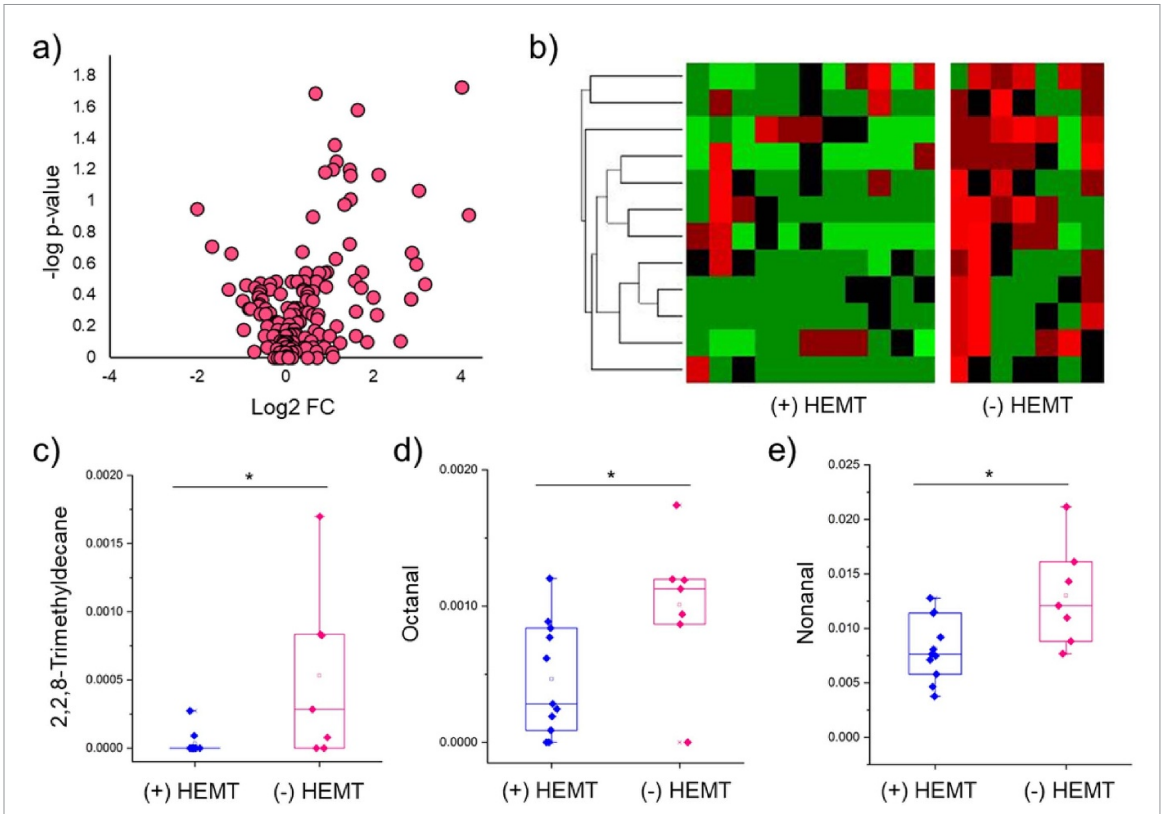

Figure 3. Analysis of VOCs by HEMT identifies some molecules significantly influenced by its use. (a) Volcano plot interpolating statistical significance and a function of fold change for comparing subjects by use of HEMT. (b) Heatmap of the VOCs identified with $p < 0.10$ and (c)–(e) box/whisker plots of the top VOCs with the highest power to discriminate subjects using HEMT.

3,7-dimethyldecane (mean normalized signal = 0.003, standard deviation = 0.003, p-value = 0.003), durene (mean normalized signal = 0.0002, standard deviation = 0.0002, p-value = 0.006), 2,4,4-Trimethyl-1,3-pentanediol 1-isobutyrate (mean normalized signal = 0.0004, standard deviation = 0.0006, p-value = 0.049) and 5-methyltridecane (mean normalized signal = 0.0005, standard deviation = 0.001, p-value = 0.049). Three out of the four VOCs were downregulated (3,7-dimethyldecane, durene, and 5-methyltridecane), while one was upregulated (2,4,4-trimethyl-1,3-pentanediol 1-isobutyrate). Because it had the highest correlation with FEV1pp and PEx, receiver operator characteristic (ROC) curves were developed to visualize and compare how 3,7-dimethyldecane and FEV1pp could predict PEx (figure 4(d)). The ROC area under the curve (AUC) was much higher for 3,7-dimethyldecane (ROC AUC = 0.91, sensitivity = 100% and specificity = 73%) when compared to FEV1pp at the time of collection (ROC AUC = 0.83, sensitivity = 86% and specificity = 82%). Of note, ΔFEV1pp had the greatest ability to distinguish PEx, and had a ROC AUC equal to 0.93 (86% sensitivity, 91% specificity), which is unsurprising because it is intrinsic to the recognition of PEx. Lastly, it should be noted that none of the statistical comparisons (BMI, age, HEMT, inhaled tobramycin, FEV1pp and PEx) had p-value < 0.05 after adjustment via the Bonferroni procedure.

4. Discussion

4.1. Analysis of exhaled VOCs and CF clinical traits

The aim of this study was to correlate exhaled VOCs with different clinical features of subjects with CF, including age, BMI, use of HEMT and/or inhaled tobramycin, FEV1pp and PEx status. Before VOC analysis, it was determined there were no significant differences in age or BMI between categorical variables of interest (HEMT, inhaled tobramycin, PEx). The only significant relationships identified were that FEV1pp was lower and ΔFEV1pp was greater in subjects with CF during PEx (figure 2). This is not a surprising correlation because FEV1pp is used for PEx prognostics and patients with PEx will have a lower FEV1pp due to decreases in pulmonary function.

Next, the team identified VOCs with significant correlations with the quantitative confounding variables including age and BMI. These VOCs are removed from this analysis because no significant differences were identified among the categorical variables, making them unlikely to contribute to clinical features of interest (figure 2). Moreover, age and BMI are confounding factors that impact the expression of VOCs in breath [39, 40]. None of these VOCs had $p < 0.05$ after implementing the Bonferroni method, and the number of VOCs found with $p < 0.05$ were not more than expected in a randomized assay, indicating VOCs were not highly impacted by these factors. After analyzing quantitative confounding variables,

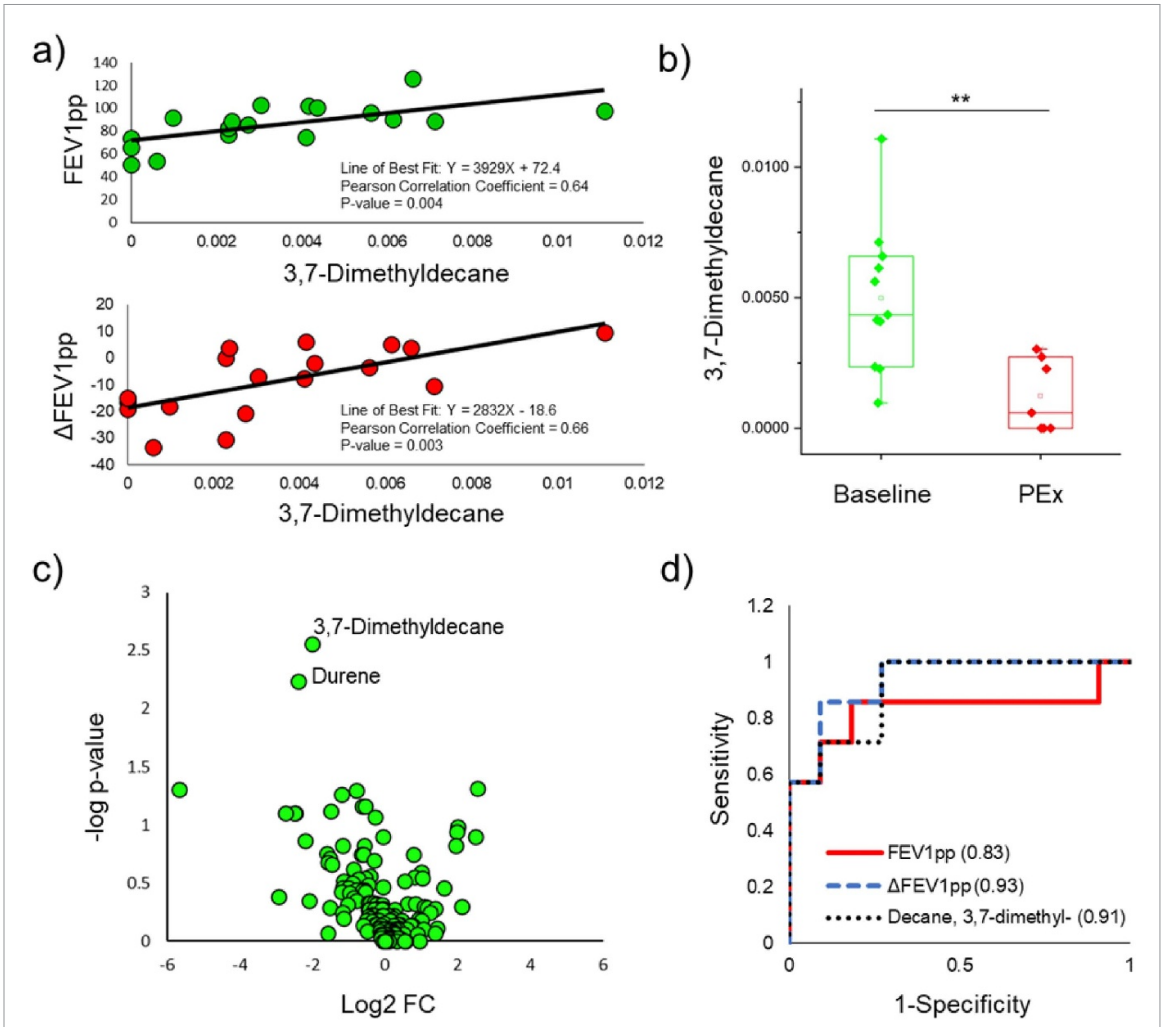

Figure 4. (a) Linear regression of 3,7-dimethyldecane shows the highest significant correlations with respect to FEV1pp. (b) Box/whisker plot of the same VOC which was the most significantly different in subject with CF experiencing PEx. (c) Volcano plot shows most VOCs are depleted during PEx. (d) Receiver operator characteristic shows that 3,7-dimethyldecane has higher accuracy (AUC 0.91) for classifying PEx relative to FEV1pp (AUC 0.83).

VOCs were probed for differences due to the use of HEMT. Four VOCs were identified to be significantly impacted (p-value < 0.05) by using HEMT (figure 3). These VOCs consisted of volatile aldehydes/saturated hydrocarbons and were identified to be upregulated in patients who were not taking HEMT. Volatile aldehydes and hydrocarbons have been previously shown to be related to increases in oxidative stress and generation of reactive oxygen species [41, 42]. This may be an indication that subjects not taking HEMT are accompanied with increases in oxidative stress. Regardless, VOC biomarkers of HEMT may not be related to PEx in this study (as many patients taking a HEMT experienced PEx, see table 1) but may be useful to monitor if patients are taking HEMT as prescribed. Even though the VOCs found to be differentially expressed by the use of HEMT did not reach p-value < 0.05 after adjustment via the Bonferroni method, 12 VOCs were identified with p-value < 0.10, which is almost 2× the number of VOCs expected at this threshold in a randomized analysis. VOCs were also analyzed for differences in patients receiving inhaled tobramycin and only one VOC was differentially expressed; however, these findings did not outperform randomized assays, indicating little to no effect of antibiotic use on VOC expression. This study specifically proposes a framework for VOC analysis in analyzing the effects of confounders and unrelated clinical variables and removing VOCs impacted by these factors. In this small sample cohort and pilot study, neither age, BMI, the use of HEMT, nor the need for inhaled tobramycin correlated significantly with lung function (FEV1pp or PEx); therefore, VOCs correlated with these clinical variables were not analyzed alongside FEV1pp or PEx.

4.2. Exhaled VOC correlations with pulmonary function

VOCs were analyzed to identify any correlations with FEV1pp, ΔFEV1pp and differences due to PEx. FEV1pp is not a confounding variable because it is an indicator of pulmonary function and is used for PEx diagnosis/prognosis [43]. The VOC with the highest Pearson Correlation Coefficient with FEV1pp

and ΔFEV1pp (3,7-dimethyldecane) was the same VOC that had the lowest p-value when comparing baseline subjects and others experiencing PEx. Because FEV1pp and PEx are intrinsically linked, they are not considered to be confounding variables and it is highly desirable to have VOC expression correlate to both factors. In both comparisons (FEV1pp and PEx), this VOC had a p-value < 0.01, indicating high significance. Three additional VOCs were identified (durene, 2,4,4-Trimethyl-1,3-pentanediol 1-isobutyrate and 5-methyltridecane) that did not correlate with FEV1pp or ΔFEV1pp but were differentially expressed in PEx. None of the VOCs had a p-value < 0.05 after adjustment using the Bonferroni approach, but two VOCs (3,7-dimethyldecane and durene) had $p \leqslant 0.006$, which is more than 4× the number of volatiles expected in a randomized assay. Moreover, strict statistical analysis through p-value adjustment may not be particularly robust and may discount VOCs differentially expressed as non-significant [44]. Even though 3,7-dimethyldecane correlated with FEV1pp, this VOC could predict PEx with higher accuracy relative to FEV1pp at the time of collection (figure 4(d)). However, ΔFEV1pp had higher classification accuracy relative to 3,7-dimethyldecane. This is not surprising since ΔFEV1pp is currently used for PEx diagnosis. Nonetheless, the use of VOCs should increase or complement the ability to stratify people with CF with PEx relative to FEV1pp in future studies. This can be accomplished by implementing machine learning and multivariate chemometric approaches in larger sample cohorts. These specific VOCs identified in this study do not directly overlap with VOCs that were previously reported as potential PEx biomarkers [37], but there are commonalities in the functional groups identified such as saturated hydrocarbons. The same study also identified xylene as a biomarker, which is interesting as it is structurally similar to durene. The origin of the VOCs identified in this study is ambiguous. These VOCs could be produced by bacteria or the alteration of cellular metabolism and other physiological processes including inflammatory pathways. Nevertheless, the downregulation of VOCs identified in PEx patients suggests that the VOCs are not produced microbially.

Avoiding unnecessary treatments and courses of antibiotics for people with CF is a top priority [45]. Differentiating respiratory symptoms that would benefit from antibiotic therapy from symptoms that will resolve spontaneously is difficult for clinicians treating people with CF, and this will become even more so for people with CF on HEMT. The decision to treat is further complicated when treatments occur over the phone, or via telehealth during the COVID-19 pandemic. The identification of VOCs that could potentially guide this treatment decision would have great potential benefit for people with CF.

4.3. Limitations

This pilot study had several limitations. Its low sample size, particularly within a fairly heterogeneous sample group (different CF genotypes, different levels of PEx, etc), is the most significant limitation. Limited number of samples decreased statistical power, and all the statistical comparisons presented did not have a p-value < 0.05 after adjustment using the Bonferroni approach. Therefore, the study is speculative in nature and spurious correlations could have been identified for the comparisons. The small number of samples also constrained the use of multivariate analysis techniques which would produce overfit models. However, its nature as a data-generating pilot study will hopefully help guide more focused future studies. It would also be of benefit to evaluate exhaled VOCs throughout the PEx process to monitor the relevant range of specific biomarkers in relation to disease severity. Also, the definition of PEx is not standardized and is commonly dependent upon the individual clinician's decision to treat. Another limitation that is global to all breath analysis studies is that VOCs in human breath can be highly variable within and between subjects in the short- and long-term. Additionally, breath samples from healthy subjects without CF were not collected in this study. Many of the VOCs identified in this study are anticipated to be detected in healthy subjects as the numerous VOCs are downregulated due to clinical traits of CF (including 3,7-dimethyldecane, durene, and 5-methyltridecane).

Finally, the sampling technique used, while validated, has limitations. Samples were collected in two different locations and breath sampling was not standardized by time, exhaled flow, or capnography but rather was collected as 'whole breath' during non-forced exhalation. This results in variations between subjects of lower and upper airway contribution to the sample and limited the ability to determine the intrapulmonary source of the measured VOCs. We recommend future studies in more focused subject populations be conducted using methods that standardize breath collection by time, volume, and capnography. Future work will additionally entail implementing machine learning (multivariate statistical analyses) to identify a biosignature of VOC biomarkers related to PEx and validating the VOCs identified in this study which correlated to the use of HEMT and PEx.

5. Conclusion

The results from this study show that exhaled VOCs are potentially related to clinical factors including age, BMI, and use of inhaled tobramycin. However, a very limited number of VOCs were identified to be correlated to these clinical variables. The VOCs detected in breath also show significant differences due to the use of HEMT, namely upregulated volatile

aldehydes (octanal and nonanal), which may be useful in the future to monitor patient adherence. Most interestingly, 3,7-dimethyldecane was identified in this study to not only significantly correlate with FEV1pp, but also has the highest ability out of all the individual VOCs to discriminate PEx. Other VOCs identified to be differentially expressed due to PEx include durene and 2,4,4-trimethyl-1,3-pentanediol 1-isobutyrate. The results presented in this study are based on a relatively small sample cohort with limited statistical power, and therefore there is a potential for identification of specious correlations. Nonetheless, the methods and approach presented in this paper may be useful to identify VOCs that correlate with confounding factors and clinical variables in other studies.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Conflict of interest

MDD is funded by NIH/NHLBI 1 PO1 HL128192, 1 PO1 HL158507-01, and Indiana CTSI UL1TR002529. MDD is a cofounder of Airbase Breathing Company and patent holder of Optate. JLS is funded by the Cystic Fibrosis Foundation 1st and 2nd Year Clinical Fellowship Grant. Other authors have no relevant COI or funding to disclose related to this work.

ORCID iDs

A P Siegel https://orcid.org/0000-0003-4112-0212
M Agarwal https://orcid.org/0000-0003-3039-0073
M D Davis https://orcid.org/0000-0001-9387-1961

References

[1] Leung D H, Heltshe S L, Borowitz D, Gelfond D, Kloster M, Heubi J E, Stalvey M, Ramsey B W and Baby O 2017 Nutrition study investigators of the cystic fibrosis foundation therapeutics development n. effects of diagnosis by newborn screening for cystic fibrosis on weight and length in the first year of life *JAMA Pediatr.* **171** 546
[2] VanDevanter D R, Kahle J S, O'Sullivan A K, Sikirica S and Hodgkins P S 2016 Cystic fibrosis in young children: a review of disease manifestation, progression, and response to early treatment *J. Cyst. Fibros.* **15** 147–57
[3] Ratjen F and Doring G 2003 Cystic fibrosis *Lancet* **361** 681–9
[4] Ferkol T, Rosenfeld M and Milla C E 2006 Cystic fibrosis pulmonary exacerbations *J. Pediatr.* **148** 259–64
[5] Liou T, Adler F, Fitzsimmons S, Cahill B, Hibbs J and Marshall B 2001 Predictive 5-year survivorship model of cystic fibrosis *Am. J. Epidemiol.* **153** 345–52
[6] Mayer-Hamblett N, Rosenfeld M, Emerson J, Goss C and Aitken M 2002 Developing cystic fibrosis lung transplant referral criteria using predictors of 2-year mortality *Am. J. Respir. Crit. Care Med.* **166** 1550–5
[7] de Boer K *et al* 2011 Exacerbation frequency and clinical outcomes in adult patients with cystic fibrosis *Thorax* **66** 680–5
[8] Konstan M *et al* 2007 Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis *J. Pediatr.* **151** 134–9 139.e1
[9] Britto M, Kotagal U, Hornung R, Atherton H, Tsevat J and Wilmott R 2002 Impact of recent pulmonary exacerbations on quality of life in patients with cystic fibrosis *Chest* **121** 64–72
[10] Lieu T, Ray G, Farmer G and Shay G 1999 The cost of medical care for patients with cystic fibrosis in a health maintenance organization *Pediatrics* **103** e72
[11] Wagener J S, Rasouliyan L, VanDevanter D R, Pasta D J, Regelmann W E, Morgan W J, Konstan M W and (Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis) 2013 Oral, inhaled, and intravenous antibiotic choice for treating pulmonary exacerbations in cystic fibrosis *Pediatr. Pulmonol.* **48** 666–73
[12] Esther C R Jr, Lin F C, Kerr A, Miller M B and Gilligan P H 2014 Respiratory viruses are associated with common respiratory pathogens in cystic fibrosis *Pediatr. Pulmonol.* **49** 926–31
[13] Aaron S D, Ramotar K, Ferris W, Vandemheen K, Saginur R, Tullis E, Haase D, Kottachchi D, St Denis M and Chan F 2004 Adult cystic fibrosis exacerbations and new strains of Pseudomonas aeruginosa *Am. J. Respir. Crit. Care Med.* **169** 811–5
[14] Goss C, Newsom S, Schildcrout J, Sheppard L and Kaufman J 2004 Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis *Am. J. Respir. Crit. Care Med.* **169** 816–21
[15] Eakin M N, Bilderback A, Boyle M P, Mogayzel P J and Riekert K A 2011 Longitudinal association between medication adherence and lung health in people with cystic fibrosis *J. Cyst. Fibros.* **10** 258–64
[16] Goss C and Burns J 2007 Exacerbations in cystic fibrosis. 1: epidemiology and pathogenesis *Thorax* **62** 360–7
[17] Carmody L A, Zhao J, Schloss P D, Petrosino J F, Murray S, Young V B, Li J Z and LiPuma J J 2013 Changes in cystic fibrosis airway microbiota at pulmonary exacerbation *Ann. Am. Thorac. Soc.* **10** 179–87
[18] Kraemer R, Deloséa N, Ballinari P, Gallati S and Crameri R 2006 Effect of allergic bronchopulmonary aspergillosis on lung function in children with cystic fibrosis *Am. J. Respir. Crit. Care Med.* **174** 1211–20
[19] Belizário J E, Faintuch J and Malpartida M G 2021 Breath biopsy and discovery of exclusive volatile organic compounds for diagnosis of infectious diseases *Front. Cell. Infect. Microbiol.* **10** 564194
[20] Saalberg Y and Wolff M 2016 VOC breath biomarkers in lung cancer *Clin. Chim. Acta* **459** 5–9
[21] Schnabel R, Fijten R, Smolinska A, Dallinga J, Boumans M-L, Stobberingh E, Boots A, Roekaerts P, Bergmans D and van Schooten F J 2015 Analysis of volatile organic compounds in exhaled breath to diagnose ventilator-associated pneumonia *Sci. Rep.* **5** 17179
[22] Phillips M, Cataneo R N, Saunders C, Hope P, Schmitt P and Wai J 2010 Volatile biomarkers in the breath of women with breast cancer *J. Breath Res.* **4** 026003
[23] Phillips M, Cataneo R N, Condos R, Erickson G A R, Greenberg J, La Bombardi V, Munawar M I and Tietje O 2007 Volatile biomarkers of pulmonary tuberculosis in the breath *Tuberculosis* **87** 44–52
[24] Janfaza S, Khorsand B, Nikkhah M and Zahiri J 2019 Digging deeper into volatile organic compounds associated with cancer *Biol. Methods Protoc.* **4** bpz014
[25] Di Gilio A *et al* 2020 Breath analysis for early detection of malignant pleural mesothelioma: volatile organic compounds (VOCs) determination and possible biochemical pathways *Cancers* **12** 1262
[26] Španěl P, Sovová K, Dryahina K, Doušová T, Dřevínek P and Smith D 2016 Do linear logistic model analyses of volatile

biomarkers in exhaled breath of cystic fibrosis patients reliably indicate Pseudomonas aeruginosa infection? *J. Breath Res.* **10** 036013
[27] Lawal O, Knobel H, Weda H, Bos L D, Nijsen T M E, Goodacre R and Fowler S J 2018 Volatile organic compound signature from co-culture of lung epithelial cell line with Pseudomonas aeruginosa *Analyst* **143** 3148–55
[28] Shestivska V, Nemec A, Dřevínek P, Sovová K, Dryahina K and Spaněl P 2011 Quantification of methyl thiocyanate in the headspace of *Pseudomonas aeruginosa* cultures and in the breath of cystic fibrosis patients by selected ion flow tube mass spectrometry *Rapid Commun. Mass Spectrom.* **25** 2459–67
[29] Angle C, Waggoner L P, Ferrando A, Haney P and Passler T 2016 Canine detection of the volatilome: a review of implications for pathogen and disease detection *Front. Vet. Sci.* **3** 47
[30] Brooks S W, Moore D R, Marzouk E B, Glenn F R and Hallock R M 2015 Canine olfaction and electronic nose detection of volatile organic compounds in the detection of cancer: a review *Cancer Invest.* **33** 411–9
[31] Van de Kant K D, van der Sande L J, Jöbsis 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** 1–23
[32] Woollam M, Wang L, Grocki P, Liu S, Siegel A P, Kalra M, Goodpaster J V, Yokota H and Agarwal M 2021 Tracking the progression of triple negative mammary tumors over time by chemometric analysis of urinary volatile organic compounds *Cancers* **13** 1462
[33] Woollam M, Teli M, Liu S, Daneshkhah A, Siegel A P, Yokota H and Agarwal M 2020 Urinary volatile terpenes analyzed by gas chromatography–mass spectrometry to monitor breast cancer treatment efficacy in mice *J. Proteome Res.* **19** 1913–22
[34] Woollam M, Teli M, Angarita-Rivera P, Liu S, Siegel A P, Yokota H and Agarwal M 2019 Detection of volatile organic compounds (VOCs) in urine via gas chromatography–mass spectrometry QTOF to differentiate between localized and metastatic models of breast cancer *Sci. Rep.* **9** 1–12
[35] Siegel A P, Daneshkhah A, Hardin D S, Shrestha S, Varahramyan K and Agarwal M 2017 Analyzing breath samples of hypoglycemic events in type 1 diabetes patients: towards developing an alternative to diabetes alert dogs *J. Breath Res.* **11** 026007
[36] Robroeks C M H H T *et al* 2010 Metabolomics of volatile organic compounds in cystic fibrosis patients and controls *Pediatr. Res.* **68** 75–80
[37] van Horck M, Smolinska A, Wesseling G, de Winter-de Groot K M, de Vreede I, Winkens B, Jöbsis Q, Dallinga J, Dompeling E and van Schooten F J 2021 Exhaled volatile organic compounds detect pulmonary exacerbations early in children with cystic fibrosis: results of a one-year observational pilot study *J. Breath Res.* **15** 026012
[38] Schumer E M, Trivedi J R, van Berkel V, Black M C, Li M, Fu X-A and Bousamra I I, M 2015 High sensitivity for lung cancer detection using analysis of exhaled carbonyl compounds *J. Thorac. Cardiovasc. Surg.* **150** 1517–24
[39] Dragonieri S, Quaranta V N, Carratu P, Ranieri T and Resta O 2016 Influence of age and gender on the profile of exhaled volatile organic compounds analyzed by an electronic nose *J. Bras. Pneumol.* **42** 143–5
[40] Cozzolino R, De Giulio B, Marena P, Martignetti A, Günther K, Lauria F, Russo P, Stocchero M and Siani A 2017 Urinary volatile organic compounds in overweight compared to normal-weight children: results from the Italian I.Family cohort *Sci. Rep.* **7** 15636
[41] Ratcliffe N, Wieczorek T, Drabińska N, Gould O, Osborne A and De Lacy Costello B 2020 A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body *J. Breath Res.* **14** 034001
[42] Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, Casalini A and Mutti A 2010 Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME–GC/MS *J. Chromatogr.* B **878** 2643–51
[43] Sanders D B, Hoffman L R, Emerson J, Gibson R L, Rosenfeld M, Redding G J and Goss C H 2010 Return of FEV1 after pulmonary exacerbation in children with cystic fibrosis *Pediatr. Pulmonol.* **45** 127–34
[44] Perneger T V 1998 What's wrong with Bonferroni adjustments *Bmj* **316** 1236–8
[45] Rowbotham N J *et al* 2018 The top 10 research priorities in cystic fibrosis developed by a partnership between people with CF and healthcare providers *Thorax* **73** 388–90