# Volatile organic compound breath signatures of children with cystic fibrosis by real-time SESI-HRMS

Ronja Weber[1,6], Naemi Haas[1,6], Astghik Baghdasaryan[1,2], Tobias Bruderer [1,3], Demet Inci[1], Srdjan Micic[1], Nathan Perkins[4], Renate Spinas[1], Renato Zenobi[3] and Alexander Moeller[1] for the Paediatric Exhalomics Group at the University Children's Hospital Zurich[5]

**Affiliations**: [1]Division of Respiratory Medicine and Childhood Research Center, University Children's Hospital Zürich, Zürich, Switzerland. [2]Joint Medical Center Arabkir, Division of Pulmonology, Yerevan, Armenia. [3]ETH Zürich, Dept of Chemistry and Applied Bioscience, Zürich, Switzerland. [4]Division of Clinical Chemistry and Biochemistry, University Children's Hospital Zürich, Zürich, Switzerland. [5]A list of the members of the Paediatric Exhalomics Group can be found at the end of this article. [6]These authors contributed equally.

**Correspondence**: Alexander Moeller, Division of Respiratory Medicine and Childhood Research Center, University Children's Hospital Zürich, Steinwiesstrasse 75, Zurich 8032, Switzerland. E-mail: alexander.moeller@kispi.uzh.ch

ABSTRACT Early pulmonary infection and inflammation result in irreversible lung damage and are major contributors to cystic fibrosis (CF)-related morbidity. An easy to apply and noninvasive assessment for the timely detection of disease-associated complications would be of high value. We aimed to detect volatile organic compound (VOC) breath signatures of children with CF by real-time secondary electrospray ionisation high-resolution mass spectrometry (SESI-HRMS).

A total of 101 children, aged 4–18 years (CF=52; healthy controls=49) and comparable for sex, body mass index and lung function were included in this prospective cross-sectional study. Exhaled air was analysed by a SESI-source linked to a high-resolution time-of-flight mass spectrometer. Mass spectra ranging from m/z 50 to 500 were recorded.

Out of 3468 m/z features, 171 were significantly different in children with CF (false discovery rate adjusted p-value of 0.05). The predictive ability (CF *versus* healthy) was assessed by using a support-vector machine classifier and showed an average accuracy (repeated cross-validation) of 72.1% (sensitivity of 77.2% and specificity of 67.7%).

This is the first study to assess entire breath profiles of children with SESI-HRMS and to extract sets of VOCs that are associated with CF. We have detected a large set of exhaled molecules that are potentially related to CF, indicating that the molecular breath of children with CF is diverse and informative.

@ERSpublications

**This is the first study to assess entire breath profiles of 101 children with SESI-HRMS. 171 molecules that are potentially related to CF were detected, indicating that the molecular breath profile of children with CF is diverse and informative.** http://bit.ly/33nIeML

**Cite this article as:** Weber R, Haas N, Baghdasaryan A, *et al.* Volatile organic compound breath signatures of children with cystic fibrosis by real-time SESI-HRMS. *ERJ Open Res* 2020; 6: 00171-2019 [https://doi.org/10.1183/23120541.00171-2019].

## Introduction

Breath contains information about processes of the human body, such as metabolism and inflammation [1, 2]. A variety of metabolites, also called volatile organic compounds (VOCs), have already been identified in exhaled breath, some of which were shown to be directly linked to diseases [3, 4]. Cystic fibrosis (CF) lung disease is a consequence of a vicious circle of early and often subclinical pulmonary infection and inflammation resulting in irreversible lung damage [5–9]. Due to its noninvasive character, breath analysis seems to be particularly attractive as a disease monitoring tool in CF.

Only a handful of studies investigated exhaled VOCs in patients with CF by mass spectrometry. In almost all previous studies, breath was collected in bags and then analysed either directly by selected ion flow tube mass spectrometry (SIFT-MS) or by thermal desorption and subsequent gas chromatography mass spectrometry (GC-MS). Two small pilot studies showed distinct patterns of a range of exhaled features in patients with CF compared to healthy controls [10, 11] detected by GC-MS. Other studies found molecules predictive for airway colonisation with various respiratory pathogens [12, 13] including *Pseudomonas aeruginosa* [14, 15]. Similar approaches reported exhaled molecules associated with CF and airway obstruction [16], acute exacerbations [17] and steroid treatment [18].

Secondary electrospray ionisation high-resolution mass spectrometry (SESI-HRMS) is a promising methodology for online breath analysis with the potential to analyse entire breath profiles. Several thousand mass-to-charge ratios (m/z features), which effectively correspond to exhaled VOCs, are detected simultaneously. Disease-specific features can subsequently be identified by bioinformatic analysis. Additionally, SESI-HRMS opens up the opportunity for compound identification of relevant molecules [19]. SESI-HRMS is a very gentle and sensitive technique that ionises volatile and even semi-volatile breath metabolites at ambient conditions [20, 21]. It can be performed in positive- and negative-ionisation mode, which increases the number and classes of compounds that can be detected. Previous studies with SESI-HRMS have identified potential exhaled biomarkers for respiratory diseases such as chronic obstructive pulmonary disease [22, 23] and obstructive sleep apnoea [24]. A recent perspective article summarised the potential of SESI-HRMS for breath biomarker discovery and the methodological issues that must be overcome in the future [25].

In a recent study, our group identified significant differences between breath profiles of adolescent and adult patients with CF and healthy age-matched controls by SESI-HRMS [26]. The present study exclusively investigates children with a methodology optimised for younger children and a larger sample size.

We hypothesised that children with CF exhale a set of disease-specific VOCs that is discriminative from healthy children. The aim of this study was a proof-of-principle for the application of online breath analysis by SESI-HRMS in children with CF and its potential to detect VOC breath signatures specific for CF. Identification of novel disease-relevant molecules could greatly improve the management of CF by providing information about the course of disease, therapy success, bacterial infections and exacerbations. Additionally, it might provide novel insights into the molecular pathophysiology of CF. Ultimately, the identification of disease-specific biomarkers in breath sets the path for rapid, noninvasive and risk-free diagnostics [2, 27, 28]. Due to the simple and noninvasive nature, molecular breath analysis is highly relevant for paediatric applications and has the potential to revolutionise diagnosis and the management of various respiratory diseases in the future.

## Methods

### Study design and participants

Children with CF and healthy children, aged between 4 and 18 years, were included in this prospective cross-sectional study. The CF population was recruited from the outpatient clinic of the University Children's Hospital Zürich, Switzerland. CF diagnosis was confirmed in all patients *via* a combination of typical clinical features, an abnormal sweat test and/or genetic testing. They were enrolled for the study during their regular follow-up visits. Healthy participants were recruited by means of flyers distributed throughout the hospital and oral communication from public schools. Measurements were planned randomly across the daytime and over the entire measurement period in order to minimise bias by circadian changes of metabolism and instrumental drift over time. Exclusion criteria included the presence of an acute infection within the last 2 weeks and the inability to perform reproducible exhalations. Patients with CF with a known colonisation with *Burkholderia cepacia*, methicillin-resistant *Staphylococcus aureus* and atypical mycobacteria, such as multidrug-resistant Gram-negative bacteria were excluded due to hygiene control. Due to the exploratory study design, sample size calculation was not feasible.

Data that were analysed for the study included breath analysis measurements as well as clinical data including lung function test, biometric data, medication and bacterial colonisation of patients. More details about the collection of clinical data are provided in the supplementary material.

The study was approved by the ethics committee of the canton of Zürich (ID 2017.00909) and participants or their legal guardians provided appropriate written consent.

### Breath analysis

Participants did not consume any edible substances or liquids (excluding water) for at least 1 h prior to the measurement. Children were required to sit comfortably on a chair or on the lap of their parents for the measurement, then instructed to take a deep inhalation and slowly exhale into the instrument through a single-use mouthpiece connected to the heated sampling line of the ionisation source. All parts were either single-use or sterilisable and the subsequent sampling line was heated to 130°C, eliminating the need for an additional bacterial filter. A measurement consisted of at least three repeatable exhalations in a row, with a minimum length of 5 s and a pressure of 5±2 mbar each. The exhaled air was analysed in real time by a novel prototype of a secondary electrospray ionisation source (Super SESI, FIT FossilionTech, Madrid, Spain) linked to a high-resolution time-of-flight mass spectrometer (Triple TOF 5600+, AB Sciex, Concord, ON, Canada). The measurement was first performed in negative-ionisation mode at a voltage of −4500 V and then repeated in positive-ionisation mode at +4500 V after a minimum 3-min rest period. Each measurement consisted of 5–6 long and constant exhalations that were recorded in series with a short break of 20 s between each breath. Mass spectra ranging from m/z 50 to 500 were recorded at an accumulation time of 1 s. More details about the breathing manoeuvre and instrumental settings are provided in the supplementary material.

### Data pre-processing

The mass spectra were resampled onto a linearly spaced m/z axis with a resolution of 0.0005 Da ($9\times10^8$ data points, 50–500 m/z range) for which peak detection was performed. The peaks arising in breath signals were selected and filtered by selecting those that appeared in at least 30% of the samples. The resulting peaks were normalised to the total ion chromatogram (TIC) for each sample and arranged into an n×k intensity matrix, where n is the number of samples and k is the number of features. See online supplementary material for more details.

### Statistical analysis

Prior to statistical analysis, adjustment for batch effects (change of the investigator team) was performed using the proposed method in JOHNSON *et al.* [29]. The identification of the significant m/z features between patients with CF and healthy participants, was performed by applying the Mann–Whitney U-test [30] on the labelled intensity matrix. The Benjamini–Hochberg procedure was applied to control for the false discovery rate (FDR) [31]. The significance level for the FDR-adjusted p-values (*i.e.* q-values) was set to 0.05.

The predictive ability (CF *versus* healthy) of collected breath samples was assessed by using a support-vector machine (SVM) classifier with linear kernel [32]. In order to avoid using all signals from breath samples for the model development, feature selection was performed on the labelled intensity matrix using the Mann–Whitney U-test in conjunction with stability selection [33]. Variable collection for stability selection was performed by applying the Mann–Whitney U-test and keeping only the features with p-values below the 0.2-quantile of all obtained p-values. The predictive performance was evaluated in a nested cross-validation [34, 35], where the outer loop of the cross-validation (10-fold cross-validation) was used to estimate the generalisation error, while the inner loop (10-fold cross-validation) was used to optimise the hyperparameters of SVM. In order to prevent feature selection bias and overfitting, feature selection had to be performed in each loop of the cross-validation (inner and outer) [35]. The process was repeated 25 times (25 times repeated 10-fold cross-validation). More details on the data analysis are provided in the supplementary material.

### Annotation of molecular formulae

The CF-specific m/z features were annotated with molecular formulae by an automated workflow based on the recommendations of KIND and FIEHN [36] with adjustments to allow water clusters and elements commonly detected with SESI. See online supplementary material for more details.

## Results

### Participant characteristics

Overall, 52 children with CF and 49 healthy children, between the age of 4 to 18 years, were included in this study. The two groups had comparable lung function and body mass index (BMI) with a similar sex distribution. Overall, 27 of the patients with CF were homozygous for the mutation F508del, whereas 6 had residual function mutations. A detailed table showing all types of mutations is included in the table S1. Other characteristics of the CF population are shown in table 1, including colonisation with bacteria and relevant medications.

**TABLE 1 Participant characteristics**

|  | Cystic fibrosis | Healthy | p-value |
|---|---|---|---|
| Age years | 11.1±3.97 | 10.9±3.59 | 0.78 |
| BMI kg·m$^{-2}$ | 17.9±3.72 | 18.3±3.13 | 0.52 |
| Males | 32 (61.5%) | 28 (57.1%) | 0.65 |
| FEV$_1$ z-score | −0.46±1.49 | −0.19±0.88 | 0.28 |
| FVC z-score | −0.12±1.22 | 0.06±0.91 | 0.42 |
| RV/TLC % | 34.3±7.72 | NA | |
| LCI | 9.41±2.08 | NA | |
| Residual CFTR function[#] | 6 (11.5%) | NA | |
| Pancreatic sufficiency[#] | 2 (3.8%) | NA | |
| *Pseudomonas aeruginosa*[#,¶] | 8 (15.4%) | NA | |
| *Staphylococcus aureus*[#,¶] | 37 (71.2%) | NA | |
| *Haemophilus influenzae*[#,¶] | 17 (32.7%) | NA | |
| *Haemophilus parainfluenzae*[#,¶] | 11 (21.2%) | NA | |
| *Aspergillus fumigatus*[#,¶] | 5 (9.6%) | NA | |
| Inhalation of salbutamol[#] | 33 (63.5%) | NA | |
| Inhalation of terbutaline[#] | 10 (19.2%) | NA | |
| Inhalation of dornase α[#] | 20 (38.5%) | NA | |
| Inhaled tobramycin[#] | 6 (11.5%) | NA | |
| Inhaled hypertonic saline 3% or 6%[#] | 50 (96.0%) | NA | |
| Oral antibiotics[#,+] | 2 (3.8%) | NA | |
| Pancreatic enzymes[#,§] | 50 (96.0%) | NA | |
| Fat-soluble vitamins[#,f] | 50 (96.0%) | NA | |
| Nasal steroid[#] | 14 (26.9%) | NA | |
| Oral ursodeoxycholic acid[#] | 21 (40.4%) | NA | |
| Oral H2 antagonist[#] | 15 (28.8%) | NA | |

Data are presented as mean±SD unless otherwise stated. BMI: body mass index; FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; RV: residual volume; TLC: total lung capacity; LCI: lung clearance index; CFTR: cystic fibrosis transmembrane conductance regulator. [#]: presented as n (% within cystic fibrosis group). [¶]: positive sputum or throat swab cultures at the time of study. [+]: co-amoxicillin, n=1; azithromycin, n=1. [§]: pancrelipase. [f]: A, D, E and K.

*Breath analysis*

Data pre-processing of breath samples resulted in 3468 m/z features across 101 subjects. The 3468×101 intensity matrix was statistically analysed in an effort to identify discriminatory m/z features between the two groups (CF *versus* controls). The Mann–Whitney U-test, together with the correction for multiple testing, revealed 171 features in exhaled breath below an FDR-adjusted p-value (q-value) of 0.05, 61 (35.7%) of which showed higher intensities in children diagnosed with CF. No m/z values were exclusively present in only one group. Table 2 shows the list of the top 20 m/z values (ordered by q-value) together with annotated molecular formulae. The corresponding box plots of the top 9 m/z features are shown in figure 1 and their first two principal components in figure 2. When attempting to predict CF from breath samples, our data processing pipeline showed an average accuracy (repeated cross-validation) of 72.1% with an average sensitivity of 77.2% and an average specificity of 67.7%.

The forced expiratory volume in 1 s (FEV$_1$) and forced vital capacity (FVC) scores of patients and controls were compared to the variation of intensities of relevant CF features. Regression analysis revealed no significant results for any of the detected 171 features (table S3).

The sample sizes of different subgroups based on intake of medication and bacterial colonisation as well as their combinations were too small to investigate the influence of these factors on the breath profiles.

## Discussion

This study investigates the largest sample size with the methodology of SESI-HRMS to date, including 101 participants. Moreover, it is the first study to assess entire breath profiles of children with SESI-HRMS and extract sets of VOC signatures associated with CF. We have detected a large set of exhaled molecules that are potentially related to CF, indicating that the molecular breath of children with CF is diverse and informative. We investigated a relatively young population without advanced CF lung disease together with a low rate of complex chronic airway infections with Gram-negative bacteria and multidrug therapy. We

| m/z[#] | p-value | FDR-adjusted p-value | Hodges–Lehmann estimator (95% CI) | Putative formula |
|---|---|---|---|---|
| −151.0247 | $1.42 \times 10^{-8}$ | $1.15 \times 10^{-5}$ | 1.785 (1.174–2.624) | $C_4H_8O_6$ |
| −75.0085 | $1.81 \times 10^{-8}$ | $1.15 \times 10^{-5}$ | 0.991 (0.692–1.557) | $C_2H_4O_3$ |
| −121.0143 | $1.95 \times 10^{-8}$ | $1.15 \times 10^{-5}$ | 1.216 (0.803–1.874) | $C_3H_6O_5$ |
| −122.0195 | $3.50 \times 10^{-8}$ | $1.54 \times 10^{-5}$ | 1.283 (0.774–1.888) | $C(C13)H_6O_5$ |
| +297.0825 | $1.07 \times 10^{-6}$ | $3.94 \times 10^{-4}$ | 1.110 (0.655–1.643) | $C_{11}H_{20}O_5S_2$ |
| +445.1200 | $1.81 \times 10^{-6}$ | $5.94 \times 10^{-4}$ | 1.197 (0.627–2.033) | $C_{16}H_{20}N_4O_{11}$ |
| −94.0260 | $2.14 \times 10^{-6}$ | $6.33 \times 10^{-4}$ | 0.960 (0.582–1.309) | $H_5N_3O_3$ |
| +359.0462 | $4.02 \times 10^{-6}$ | $1.07 \times 10^{-3}$ | 0.819 (0.502–1.289) | $C_{11}H_{18}O_9S_2$ |
| +445.0985 | $4.43 \times 10^{-6}$ | $1.07 \times 10^{-3}$ | 1.162 (0.584–1.968) | $C_{19}H_{24}O_8S_2$ |
| −93.0195 | $4.73 \times 10^{-6}$ | $1.07 \times 10^{-3}$ | 1.074 (0.637–1.623) | $C_2H_6O_4$ |
| +357.0490 | $7.19 \times 10^{-6}$ | $1.52 \times 10^{-3}$ | 1.014 (0.561–1.615) | $C_{11}H_{16}O_{11}S$ |
| +447.1420 | $9.56 \times 10^{-6}$ | $1.88 \times 10^{-3}$ | 0.932 (0.476–1.541) | $C_9H_{26}N_4O_{16}$ |
| +332.1202 | $1.19 \times 10^{-5}$ | $1.97 \times 10^{-3}$ | 1.153 (0.639–1.754) | $C_{11}H_{25}NO_6S_2$ |
| −105.0188 | $1.19 \times 10^{-5}$ | $1.97 \times 10^{-3}$ | 0.917 (0.506–1.364) | $C_3H_6O_4$ |
| +429.0880 | $1.27 \times 10^{-5}$ | $1.97 \times 10^{-3}$ | 0.922 (0.507–1.472) | $C_{15}H_{24}O_{10}S_2$ |
| +445.1483 | $1.27 \times 10^{-5}$ | $1.97 \times 10^{-3}$ | 0.809 (0.457–1.338) | $C_{28}H_{20}N_4S$ |
| +188.1645 | $1.53 \times 10^{-5}$ | $2.25 \times 10^{-3}$ | −0.799 (−1.153−−0.443) | $C_{10}H_{21}NO_2$ |
| +447.0983 | $2.34 \times 10^{-5}$ | $3.29 \times 10^{-3}$ | 0.889 (0.464–1.515) | $C_{28}H_{18}N_2S_2$ |
| +299.0797 | $2.48 \times 10^{-5}$ | $3.33 \times 10^{-3}$ | 0.930 (0.548–1.492) | $C_{10}H_{18}O_8S$ |
| +359.0285 | $2.97 \times 10^{-5}$ | $3.82 \times 10^{-3}$ | 0.872 (0.463–1.436) | $C_{10}H_{14}O_{12}S$ |

TABLE 2 20 most significant mass/charge (m/z) features related to cystic fibrosis and their putative molecular formulae ordered by q-value

The entire list including all significant features and all annotated molecular formulae is provided in the table S2, including nine features that could not be assigned with a molecular formula. The intensities were scaled for better readability using distance to median divided by the median absolute deviation. FDR: false discovery rate. [#]: + indicates positive-ionisation mode and − indicates negative-ionisation mode.

therefore assume that the results of this explorative study may reflect the complex metabolic processes of the disease and its interplay with the airway microbiology, inflammation and drug metabolism.

As we were the first group to translate the methodology of SESI-HRMS into paediatrics we needed to adapt the breath sample collection to a relatively young population. By our definition, a successful measurement consisted of a sequence of at least three reproducible, long and continuous exhalations. Young children had difficulties to keep a constant pressure across the entire exhalation and to exhale for a sufficient amount of time. Nevertheless, we were able to obtain reproducible measurements from children aged 4 years, which was therefore determined to be the minimum age for participants. Although the time of exhalation was shorter than that in older children, the exhalations were still repeatable with a low intra-subject variability and had a stable plateau phase. As an example, figure 3 shows the time trace of the TIC as well as one CF-specific m/z feature 151.0247 in a 4.5-year-old and a 13-year-old patient. Nevertheless, a successful measurement was not only dependent on age, but also on the cooperation and concentration of the children and their parents.

We managed to extract breath profiles consisting of molecules discriminative of the disease state. As illustrated in the box plots of nine exemplary features (q-value<0.05), normalised intensities show relatively distinguishable values between the two groups. Figure 4 shows extracted time traces of three VOC features of a patient with CF and a healthy child of the same age (13 years old). The visible difference in absolute intensity serves as an example that our method is capable of detecting clear differences between healthy and diseased probands.

The assessment of the predictive power of breath samples resulted in an average accuracy of 72.1%. We hypothesise that the accuracy score was not surprising considering the general heterogeneity, young age and remarkable health of the investigated CF population. This, we claim, is also visible in the box plots of the most significant features (figure 1) and their first two principal components (figure 2). Firstly, the variation in intensity of the features related to patients with CF was generally greater in the patient group than the healthy one, representing the large heterogeneity of the disease. Secondly, the young age and excellent health of the patients with CF result in a slight overlap of the intensities of CF-specific features between some patients and healthy controls. Nonetheless, we could show that entire breath profiles of children diagnosed with CF carry significant information related to their condition.
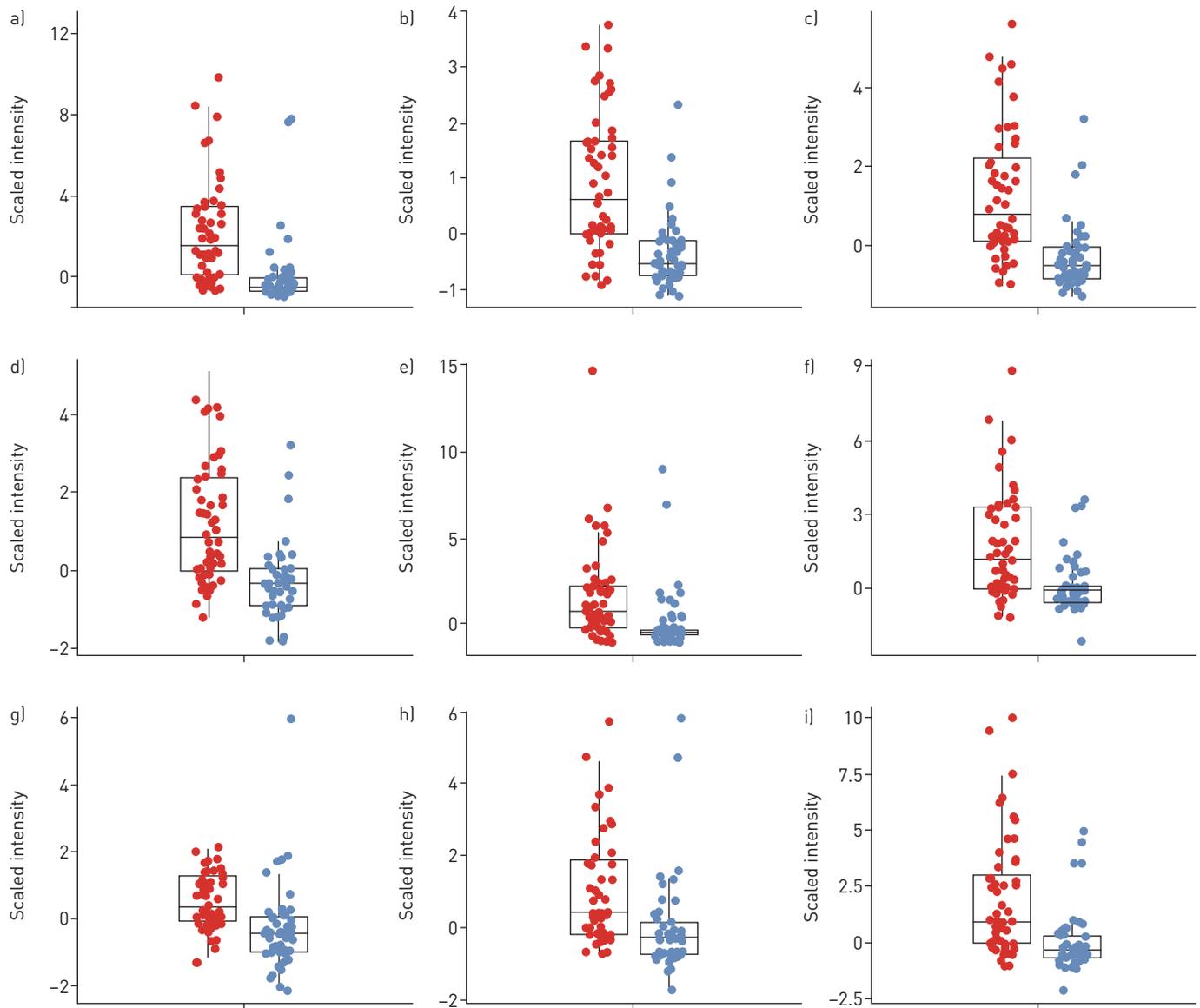
FIGURE 1 Box plots of the nine most significant cystic fibrosis (CF) features. Intensities were scaled to median absolute deviation for better readability. The features were selected by applying the Mann–Whitney U-test followed by the multiple testing correction and are ordered by their q-values. Patients with CF are shown in red, healthy controls are in blue. Features marked with a "+" sign were detected in positive- and those with a "−" sign in negative-ionisation mode. a) Mass/charge ratio (m/z)=−151.0247 (false discover rate (FDR)-adjusted p=1.2×10$^{-5}$); b) m/z=−75.0085 (FDR-adjusted p=1.2×10$^{-5}$); c) m/z=−121.0143 (FDR-adjusted p=1.2×10$^{-5}$); d) m/z=−122.0195 (FDR-adjusted p=1.5×10$^{-5}$); e) m/z=+297.0825 (FDR-adjusted p=0.00039); f) m/z=+445.0985 (FDR-adjusted p=0.0011); g) m/z=−94.0260 (FDR-adjusted p=0.00063); h) m/z=+359.0462 (FDR-adjusted p=0.0011); i) m/z=+445.1200 (FDR-adjusted p=0.00059). FDR: false discovery rate.

A stricter feature selection (*e.g.* lower p-value threshold) or even a transformation of the features space into some other space (*e.g.* principal component analysis (PCA)) could have led to a higher average accuracy of the data pipeline. However, all these measures are to be interpreted only within the scope of the current study. Consequently, any real-world interpretation can only be done with a separate external validation dataset, which we intend to create in the upcoming validation study.

The number of patients (n=52) together with the variety of patient-specific variables did not allow an extraction of breath profiles that are relevant for in-depth phenotyping of the CF population. We investigated effects of age, sex, BMI and most importantly, medication intake (including inhalation of salbutamol, inhaled corticosteroids, tobramycin, rhDNAse; oral antibiotics, pancreatic enzymes, ursodeoxycholic acid) and did not find any systematic and significant influences on the features in our statistical analyses. At this stage we cannot exclude the fact that some of the features are related to exposure to medication, either due to small numbers (*e.g.* use of inhaled tobramycin in only six (11%) children with CF) or that the drug was used in almost all patients (*e.g.* pancreatic enzymes in 96% of the patients).
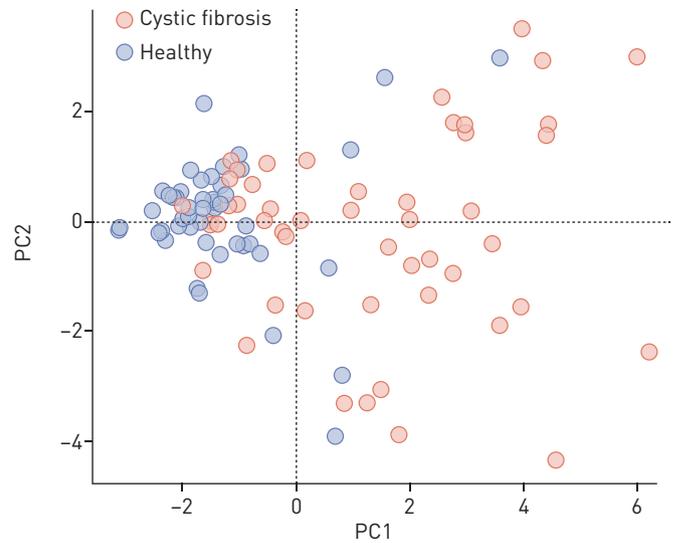
FIGURE 2 Plot of the first two principal components (PCs) of the nine most significant cystic fibrosis-specific features (see figure 1).

Due to the large heterogeneity of the disease itself, intake of many medications and the inhomogeneity of bacterial colonisation, a larger sample size with evenly distributed groups is needed for meaningful phenotyping. However, this was not the aim of the present study. Therefore, we did not attempt to identify features indicative for specific infections, as for example with *Pseudomonas aeruginosa*. Nevertheless, the timely and noninvasive detection of bacterial infection is especially attractive for breath analysis in CF, since it would be highly relevant for clinical application and we plan to investigate the potential of SESI-HRMS to detect *Pseudomonas* and other infection in breath in a subsequent prospective study. A previous study with the same methodology detected the medication salbutamol in breath [37]. Since all included patients with CF have taken a variety of inhalatory and noninhalatory medications, it is to be expected that the extracted VOC breath signature contains some molecules that are related to medication (or bacterial infection) and not the disease itself. Since our sample size was not large enough to determine the influence of these parameters by statistical means, identification of the compounds and further studies with more participants will provide more insights.

Some limitations should be noted. First, we did not measure background air samples. Although these may affect the results to some degree, all measurements were performed in a dedicated mass spectrometry laboratory with standardised room conditions and we expect background noise to be relatively stable over the study. In addition, both patients with CF and healthy controls were measured under the same conditions. We therefore expect that measurements of both groups were similarly affected by room conditions and will therefore be less relevant to group differences. We did not use an inspiratory VOC filter and did not continuously monitor the molecular composition of room air. Therefore, the presence of VOCs in the inhaled room air might have influenced detected features.

Further, to this stage, we present a broad comparison of breath profiles between children with CF and healthy controls but we do not present specific compounds. The identification of the compounds is a complex and elaborate process. This process is ongoing and the identified components will be evaluated in an upcoming validation study. We were interested in a wide spread of different ages to assess feasibility on the one hand and the whole spectrum of disease severity on the other hand. However, our CF group is a relatively stable group with similar $FEV_1$ and BMI values to the healthy controls, which limits the identification of breath profiles that may be associated with severity. The variability of the compounds seems larger in the CF group (as apparent in the box plots in figure 1). Although we did not find clear statistically significant correlations between lung function and age with the breath compounds at this stage, we hypothesise that the wider spread may be due to the variable ages and severity of lung disease.

This was the first study conducted with a new prototype SESI ion source (Super SESI, Fossil Ion Technology) coupled to a time-of-flight mass spectrometer. Not surprisingly, due to the novelty of the ion source, a few technical issues arose over time. For example, a progressive decrease of electrospray current and contaminations of the electrospray fluid. The mentioned issues did not influence our results, since we excluded almost all measurements that were recorded during the period of voltage shift and then continued the study once the physical defect was repaired. As a consequence, we started to continuously monitor the current technical state of the instrument. In order to avoid such problems in future studies,
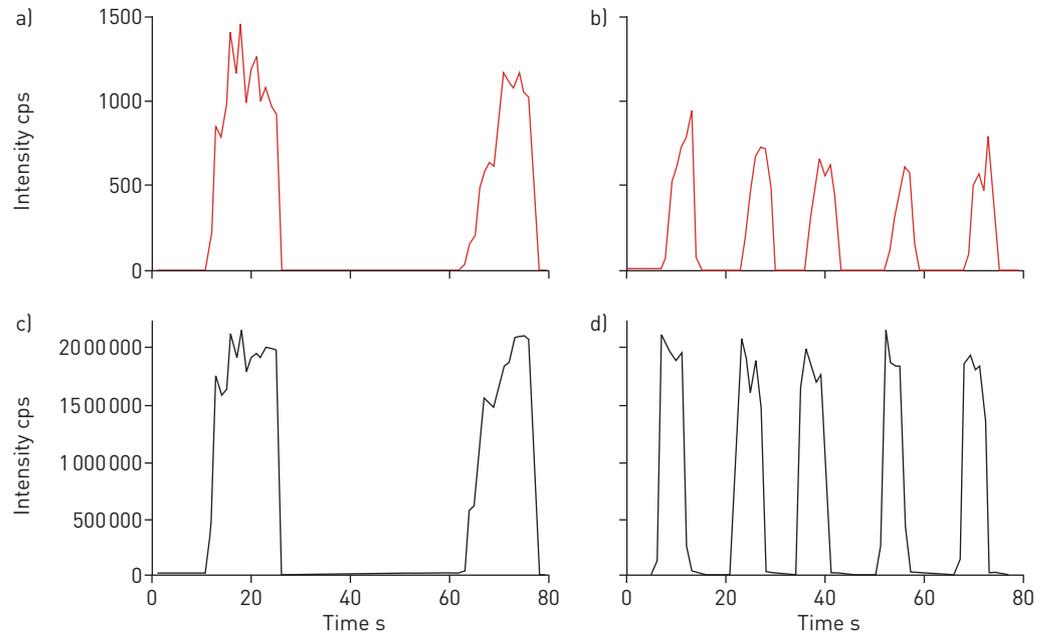
FIGURE 3 a and b) Extracted time traces of the most significant mass/charge ratio (151.0247) and c and d) total ion chromatography (TIC) recorded in negative ionisation mode from a and c) a 13-year-old and b and d) a 4.5-year-old patient with cystic fibrosis. The recorded signal exhibits good reproducibility for both age groups across several consecutive exhalations. As expected, the time of exhalation, which is clearly visible in the time trace of the TIC, was longer in older children. cps: counts per second.

there is a need for effective and robust quality control criteria in order to detect technical problems quickly. In order to further improve the methodology SESI-HRMS, the technical handling, conduction of measurements and data analysis needs standardisation among different research groups.

Interestingly, there was no overlap between our detected VOC signatures and those identified in a previous study with SESI-MS [26]. However, we did investigate a different patient group, consisting of mostly young and relatively healthy children. Additionally, we used a novel and improved ion source [25] with inert surfaces and improved cleaning capabilities. Lastly, we adjusted the HRMS setting to reduce fragmentation of detected m/z features by setting the collision gas to collision-activated dissociation (CAD)=0, compared to the standard setting of CAD=6 during the previous study. These differences of the instrumental set up make it hard to compare the results of the two studies. There was also no obvious overlap between our detected CF-specific m/z features and compounds found by other studies in patients with CF using different mass spectrometric methodologies. However, thorough comparison of studies will only be possible as soon as our features are unambiguously identified. Currently, we are able to annotate 95% of the 171 features from patients with CF (m/z features) with a putative molecular formula, in contrast to the first CF study from our group [26] where only 37% of 49 features could be assigned with a formula. The previously reported compound identification workflow [38], which was based on the analysis of exhaled breath condensates showed limited applicability for the present study. Therefore, we are currently developing a compound identification approach with direct, online breath analysis including signal-boosting and pre-separation steps. This should likely allow a thorough identification of CF-specific features in the subsequent validation study. Since there are limited databases for the matching of compounds detected by SESI-HRMS available and interpretation of mixed-fragment MS-MS spectra without pre-separation is challenging, compound identification is a time consuming and complex process However, identification is required to get insights into the molecular pathophysiology of CF. We hypothesise that we will identify molecules reflecting bacterial metabolism and that at least a part of these features will be different in patients with distinct bacterial colonisation of their airways and that relevant features reflect oxidative stress and neutrophilic airways inflammation. There is no doubt that the microbiome occurs in the airways but also the gut impacts on exhaled VOC's and we are sure that some of the differences we found may be explained by differences in the microbiome.

For future studies, our main goal is to adopt the SESI-HRMS methodology for the phenotyping of patients with CF. We are extremely interested in the early detection of bacterial infection, including *Pseudomonas aeruginosa* and acute exacerbations in breath. Although normal at birth, infants with CF develop an
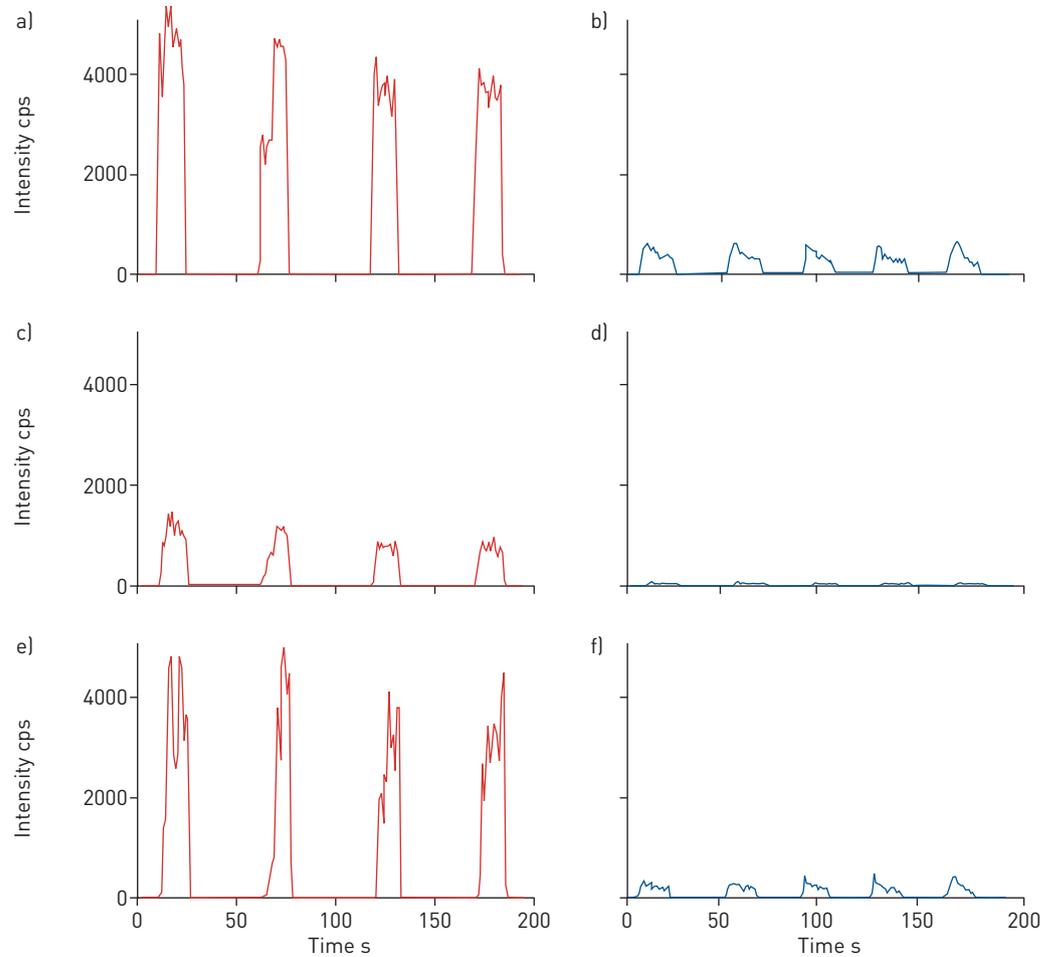
FIGURE 4 Time traces of three cystic fibrosis (CF)-specific mass/charge ratio values (a and b) 121.0143, c and d) 151.0247 and e and f) 93.0195) in negative-ionisation mode compared between a, c and e) a child diagnosed with CF and b, d and f) a healthy child of the same age (13 years). cps: counts per second.

abnormal airway microbiome during the first year of life [39]. Their airways are rapidly colonised by pathogens, including *Staphylococcus aureus* and *Haemophilus influenzae* followed by predominance of Gram-negative bacilli such as *Pseudomonas aeruginosa* [40]. Chronic infection and colonisation result in a deteriorating vicious circle and contribute substantially to disease progression and CF-related mortality [5–9]. Early detection of airway infection is crucial to start antimicrobial therapy, increase survival and reduce cross-infections. An easy to apply and noninvasive assessment of disease activity and disease-associated complications, such as airway infection, would be of high value in the management of patients with CF. Since we assume that medication could have an impact on breath signal, we will address the impact of various medications on breath composition in future studies. We further plan to investigate whether changes in the CF breath profile are related to disease progression, treatment success and other factors that are helpful for providing the best therapy and care for children with CF.

### Conclusions

The SESI-HRMS methodology is capable of detecting breath signatures in children with CF. The identification and monitoring of CF-specific VOCs in breath is highly relevant and promising for various clinical applications, such as early detection of bacterial infection, exacerbations and therapy monitoring.

Paediatric Exhalomics Group: Astghik Baghdasaryan, Christoph Berger, Christian Bieli, Tobias Bruderer, Naemi Haas, Martin Hersberger, Katharina Heschl, Demet Inci, Andreas Jung, Malcolm Kohler, Srdjan Micic, Alexander Moeller, Simona Müller, Nathan Perkins, Tina Schürmann, Florian Singer, Renate Spinas, Bettina Streckenbach, Jakob Usemann, Ronja Weber and Renato Zenobi.

## References

1    Popov TA. Human exhaled breath analysis. *Ann Allergy Asthma Immunol* 2011; 106: 451–456.
2    Amann A, Costello Bde L, Miekisch W, *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J Breath Res* 2014; 8: 034001.
3    de Lacy Costello B, Amann A, Al-Kateb H, *et al.* A review of the volatiles from the healthy human body. *J Breath Res* 2014; 8: 014001.
4    Risby TH, Solga SF. Current status of clinical breath analysis. *Applied Physics B* 2006; 85: 421–426.
5    Khan TZ, Wagener JS, Bost T, *et al.* Early pulmonary inflammation in infants with cystic fibrosis. *Am J Respir Crit Care Med* 1995; 151: 1075–1082.
6    Schogler A, Stokes AB, Casaulta C, *et al.* Interferon response of the cystic fibrosis bronchial epithelium to major and minor group rhinovirus infection. *J Cyst Fibros* 2016; 15: 332–339.
7    Hartl D, Gaggar A, Bruscia E, *et al.* Innate immunity in cystic fibrosis lung disease. *J Cyst Fibros* 2012; 11: 363–382.
8    Armstrong DS, Hook SM, Jamsen KM, *et al.* Lower airway inflammation in infants with cystic fibrosis detected by newborn screening. *Pediatr Pulmonol* 2005; 40: 500–510.
9    O'Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet* 2009; 373: 1891–1904.
10   Robroeks CM, van Berkel JJ, Dallinga JW, *et al.* Metabolomics of volatile organic compounds in cystic fibrosis patients and controls. *Pediatr Res* 2010; 68: 75–80.
11   Kramer R, Sauer-Heilborn A, Welte T, *et al.* A rapid method for breath analysis in cystic fibrosis patients. *Eur J Clin Microbiol Infect Dis* 2015; 34: 745–751.
12   Bos LD, Meinardi S, Blake D, *et al.* Bacteria in the airways of patients with cystic fibrosis are genetically capable of producing VOCs in breath. *J Breath Res* 2016; 10: 047103.
13   Neerincx AH, Geurts BP, van Loon J, *et al.* Detection of *Staphylococcus aureus* in cystic fibrosis patients using breath VOC profiles. *J Breath Res* 2016; 10: 046014.
14   Shestivska V, Nemec A, Drevinek P, *et al.* 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* 2011; 25: 2459–2467.
15   Gilchrist F, Jones A, Webb K, *et al.* The sensitivity and specificity of *Pseudomonas aeruginosa* detection using hydrogen cyanide concentration in exhaled breath – The SPACE study. *Eur Respir J* 2014; 44: Suppl. 58, 3444.
16   Barker M, Hengst M, Schmid J, *et al.* Volatile organic compounds in the exhaled breath of young patients with cystic fibrosis. *Eur Respir J* 2006; 27: 929–936.
17   McGrath L, Patrick R, Mallon P, *et al.* Breath isoprene during acute respiratory exacerbation in cystic fibrosis. *Eur Respir J* 2000; 16: 1065–1069.
18   Paredi P, Kharitonov SA, Leak D, *et al.* Exhaled ethane is elevated in cystic fibrosis and correlates with carbon monoxide levels and airway obstruction. *Am J Respir Crit Care Med* 2000; 161: 1247–1251.
19   Gaugg MT, Gomez DG, Barrios-Collado C, *et al.* Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high-resolution mass spectrometry – a pilot study on tobacco smokers. *J Breath Res* 2016; 10: 016010.
20   Martínez-Lozano P, de la Mora JF. Electrospray ionization of volatiles in breath. *Int J Mass Spectrom* 2007; 265: 68–72.
21   Gaugg MT. On-line breath metabolomics in respiratory diseases using secondary electrospray ionization-mass spectrometry. *Chimia* 2018; 72: 184–188.
22   Martinez-Lozano Sinues P, Meier L, Berchtold C, *et al.* Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease. *Respiration* 2014; 87: 301–310.
23   Gaugg MT, Nussbaumer-Ochsner Y, Bregy L, *et al.* Real-time breath analysis reveals specific metabolic signatures of COPD exacerbations. *Chest* 2019; 156: 269–276.
24   Schwarz EI, Martinez-Lozano Sinues P, Bregy L, *et al.* Effects of CPAP therapy withdrawal on exhaled breath pattern in obstructive sleep apnoea. *Thorax* 2016; 71: 110–117.
25   Singh KD, Del Miguel GV, Gaugg MT, *et al.* Translating secondary electrospray ionization-high-resolution mass spectrometry to the clinical environment. *J Breath Res* 2018; 12: 027113.
26   Gaisl T, Bregy L, Stebler N, *et al.* Real-time exhaled breath analysis in patients with cystic fibrosis and controls. *J Breath Res* 2018; 12: 036013.
27   Martinez-Lozano Sinues P, Zenobi R, Kohler M. Analysis of the exhalome: a diagnostic tool of the future. *Chest* 2013; 144: 746–749.
28   Amann A, Miekisch W, Schubert J, *et al.* Analysis of exhaled breath for disease detection. *Annu Rev Anal Chem (Palo Alto Calif)* 2014; 7: 455–482.
29   Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8: 118–127.
30   Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist* 1947; 18: 50–60.
31   Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 1995; 57: 289–300.
32   Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; 20: 273–297.
33   Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol* 2010; 72: 417–473.

34  Iizuka N, Oka M, Yamada-Okabe H, *et al.* Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet* 2003; 361: 923–929.

35  Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006; 7: 91.

36  Kind T, Fiehn O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 2007; 8: 105.

37  Gaugg MT, Engler A, Nussbaumer-Ochsner Y, *et al.* Metabolic effects of inhaled salbutamol determined by exhaled breath analysis. *J Breath Res* 2017; 11: 046004.

38  Gaugg MT, Bruderer T, Nowak N, *et al.* Mass-spectrometric detection of omega-oxidation products of aliphatic fatty acids in exhaled breath. *Anal Chem* 2017; 89: 10329–10334.

39  Mika M, Korten I, Qi W, *et al.* The nasal microbiota in infants with cystic fibrosis in the first year of life: a prospective cohort study. *Lancet Respir Med* 2016; 4: 627–635.

40  Rosenfeld M, Ramsey BW, Gibson RL. *Pseudomonas* acquisition in young patients with cystic fibrosis: pathophysiology, diagnosis and management. *Curr Opin Pulm Med* 2003; 9: 492–497.

**"Volatile organic compound breath signatures of children with cystic fibrosis by real-time SESI-HRMS". Ronja Weber, Naemi Haas, Astghik Baghdasaryan, Tobias Bruderer, Demet Inci, Srdjan Micic, Nathan Perkins, Renate Spinas, Renato Zenobi and Alexander Moeller for the Paediatric Exhalomics Group at the University Children's Hospital Zurich. *ERJ Open Res* 2020; 6: 00171-2019.**

In the originally published version of this article, it was not indicated that Ronja Weber and Naemi Haas contributed equally. This has been corrected in the article itself.