

PAPER • OPEN ACCESS

Volatile organic compounds in a headspace sampling system and asthmatics sputum samples

To cite this article: Rosa Peltrini *et al* 2021 *J. Breath Res.* **15** 027102

View the [article online](#) for updates and enhancements.



NEW BREATH BIOPSY PRODUCTS

NEW FEATURES | NEW LOOK

SAME WORLD-LEADING
BREATH RESEARCH PLATFORM

VIEW OUR NEW RANGE

owlstonemedical.com





PAPER

OPEN ACCESS

RECEIVED

7 September 2020

REVISED

1 November 2020

ACCEPTED FOR PUBLICATION

23 November 2020

PUBLISHED

11 January 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Volatile organic compounds in a headspace sampling system and asthmatics sputum samples

Rosa Peltrini¹ , Rebecca L Cordell², Wadah Ibrahim^{1,3}, Michael J Wilde², Dahlia Salman⁴ , Amisha Singapuri^{1,3}, Beverley Hargadon^{1,3}, Christopher E Brightling^{1,3}, C L Paul Thomas⁴ , Paul S Monks², Salman Siddiqui^{1,3} and on behalf of the EMBER consortium⁵

¹ University of Leicester, Department of Respiratory Sciences, LE1 7RH Leicester, United Kingdom

² University of Leicester, Department of Chemistry, LE1 7RH Leicester, United Kingdom

³ Leicester NIHR Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Groby Road, LE3 9QP Leicester, United Kingdom

⁴ Department of Chemistry, Loughborough University, Loughborough LE11 3TU, United Kingdom

⁵ See acknowledgments for contributors

E-mail: ss338@leicester.ac.uk

Keywords: volatile organic compounds, background signal, headspace analysis, severe asthma, sputum

Abstract

The headspace of a biological sample contains exogenous volatile organic compounds (VOCs) present within the sampling environment which represent the background signal. This study aimed to characterise the background signal generated from a headspace sampling system in a clinical site, to evaluate intra- and inter-day variation of background VOC and to understand the impact of a sample itself upon commonly reported background VOC using sputum headspace samples from severe asthmatics. The headspace, in absence of a biological sample, was collected hourly from 11am to 3pm within a day (time of clinical samples acquisition), and from Monday to Friday in a week, and analysed by thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). Chemometric analysis identified 1120 features, 37 of which were present in at least the 80% of all the samples. The analyses of intra- and inter-day background variations were performed on 13 of the most abundant features, ubiquitously present in headspace samples. The concentration ratios relative to background were reported for the selected abundant VOC in 36 asthmatic sputum samples, acquired from 36 stable severe asthma patients recruited at Glenfield Hospital, Leicester, UK. The results identified no significant intra- or inter-day variations in compounds levels and no systematic bias of z-scores, with the exclusion of benzothiazole, whose abundance increased linearly between 11am and 3pm with a maximal intra-day fold change of 2.13. Many of the identified background features are reported in literature as components of headspace of biological samples and are considered potential biomarkers for several diseases. The selected background features were identified in headspace of all severe asthma sputum samples, albeit with varying levels of enrichment relative to background. Our observations support the need to consider the background signal derived from the headspace sampling system when developing and validating headspace biomarker signatures using clinical samples.

1. Introduction

The extraction and the analysis of the headspace of a biological sample is an approach mostly used to detect possible biomarkers of physiological and pathological conditions affecting the sample [1–3]. The headspace of a sample consists of the vapour phase within a sealed container in which the sample is located. The vapour is generally concentrated above the sample material at the top of the container, hence the word

‘head-space’ [4]. The headspace of a biological sample contains both volatile and semi-volatile organic compounds (VOCs/SVOCs) which represent the vapour fraction of the sample, mirroring its composition and biophysical properties [5]; therefore physiological or pathological alterations of a biological sample may be reflected in qualitative and/or quantitative VOC variations in its headspace.

Headspace analysis of cell cultures (such as inflammatory cells [6] small airway epithelial cells

[7], sputum samples (as source microbes) [8] and of microbial cultures [9]), has been used, together with the analysis of volatile metabolites in exhaled breath (breathomics) [10, 11], as a non-invasive methodology to identify potential airborne biomarkers linked to different airway pathological conditions and in order to stratify different patients' phenotypes.

VOCs present in headspace of biological samples and in breath might originate from biological metabolic processes of endogenous constituents. For example acetone resulting from cellular ketogenesis [12] and is considered a systemic volatile component [12, 13] which is detected in breath and urine headspace [14]. Headspace of biological samples can also contain VOCs originating from the catabolism of external components to which individuals are exposed [15]. For example exogenous VOCs can be inhaled, ingested or parenterally absorbed, eventually metabolised and released in biological samples such as serum, plasma, urine, sputum and breath, which are routinely used in clinical metabolomics profiling studies. For example the presence of hydroxylated polycyclic aromatic hydrocarbons, such as naphthalene, phenanthrene and pyrene, was detected in the headspace of urine samples from workers who were exposed to hot asphalt [16]. The headspace of a biological sample can also contain exogenous VOCs present in the room air or VOCs that have originated from the headspace collection system. These exogenous VOCs represent part of the background signal affecting the headspace composition. The background signal can vary due to several factors such as ambient temperature and humidity of the room in which the headspace is collected, meteorological conditions [17], materials involved in the sampling process and the timing of headspace collection as demonstrated by Boshier *et al*, who identified that in the ambient air at three different clinical sites, there were higher levels of ethanol in the morning compared to the evening, and increased levels of ammonia at the end of the day compared to the morning [18]. Some of the VOCs which are reported in literature as exogenous VOCs present within clinical environment, are also reported as potential biomarkers for several disease conditions. Toluene and isopropanol, for example, are reported as common volatile components of the clinical environment [19] but are also found among the predominant VOCs in breath of asthmatic patients [20] and in headspace of faeces samples of patients with non-alcoholic fatty liver disease [21]. These observations suggest that the background VOCs and their variations could affect, for instance, the signal of a targeted VOC in the headspace of a biological sample, and support the need for a better understanding of the background signal in headspace of biological samples during the development and validation of a panel of volatile biomarkers.

With this study we aimed (a) to identify background VOCs generated within a headspace sampling

system developed for collecting headspace from sputum samples, (b) to evaluate the intra-day and inter-days background variation of common and abundantly identified features and (c) to estimate the contribution of the volatile background signal generated from the sampling system into headspace of sputum samples acquired from a cohort of stable severe asthma patients.

2. Materials and methods

2.1. Study design

The study was performed in the month of October 2019 at the Glenfield Hospital, Leicester (UK) in a clinical research facility testing room where lighting, temperature and humidity levels were constantly monitored and calibrated.

2.1.1. Background VOC variation study

The headspace collection in absence of a biological sample (blank sample)—which contained just the background VOCs generated from the headspace sampling system—was performed using a headspace sampling system consisting of a small vessel, an empty Petri dish and a supply of pure air. The headspace samples collected in absence of a biological specimens and the VOC found in them, were reported in this study respectively as background headspace and background VOCs.

For the headspace background signal intra-day variation study, the headspace from a blank sample was collected—using the method subsequently described—hourly from 11am to 3pm. The time intervals were chosen based on the period within which clinical research samples (e.g. sputum, urine, and blood) are usually collected and processed. The experiments were repeated in quadruplicate (per time interval), generating 20 headspace samples in order to estimate the intra-day background signal variations. For the inter-days variations study of the background headspace signal, the headspace from a blank sample was collected every day for a week, from Monday to Friday between 11am and 3pm. The experiment was repeated in triplicate, generating 15 headspace samples. The same analytical process consisting of TD-GC-MS analysis was applied to the headspace samples collected for both the intra- and inter-days background variation studies. The TD-GC-MS data of both the studies were integrated, deconvolved and aligned (as subsequently described) to generate a matrix of 1120 features. A feature represents a peak in the chromatogram that satisfies the adopted integration parameters (s/n, peak area, peak height etc.).

The 1120 identified features were used to (a) characterise the background volatile signature generated from the headspace sampling system represented by the features present in at least the 80% of all the headspace samples (for both the intra and inter-days variation studies); and (b) to identify features

ubiquitously present across all the 35 samples which were used for statistical analysis of background signal intra-day and inter-days variation in headspace samples.

2.1.2. Sputum headspace asthma study

A separate and subsequent study was conducted to evaluate the occurrence of the most abundant and ubiquitous present background features in headspace of sputum samples from severe asthmatic patients. The cohort of severe asthmatics consisted of an adult population with an age (median (Q1:Q3)) of 62 (57:70.25) years; 14 females and 22 males with a post bronchodilator FEV₁ (Forced Expiratory Volume in the first second of the forced breath)% predicted of 69.75 (52.50:87.60) and a post-bronchodilator FEV₁/FVC (forced vital capacity)% predicted of 87.78 (74.4:97.7). According to the Global Initiative for Asthma (GINA) [22] treatment strategy for adults, 20 patients were classified as GINA 5 and 16 as GINA 4; 14 of the GINA 5 patients were also receiving anti-IL5 therapy (Mepolizumab).

The sputum samples were spontaneously produced and acquired at least 6 weeks post exacerbation in the asthma cohort. The sputa were collected in the clinical research facility testing room where the background experiments were conducted and where the headspace was subsequently sampled. Headspace from a blank sample was collected for each sputum headspace as control. Based on the results of background headspace variation study (within a day and between different days) the sampling time for sputum samples and the respective controls was between 11am and 3pm (no specific time point selection was needed). The 36 sputum samples were collected from Monday to Friday over a period time of 12 months; the headspace from sputum samples and their controls were collected immediately after the sputum production and then analysed by TD-GC-MS.

The most abundant background features identified in the previous study were identified as well in control headspace samples and compared with those found in headspace of sputum samples.

2.2. Method of headspace collection

The sampling system and the method used to collect headspace in presence and absence of the sputum sample were the same and were based on the generation of a dynamic headspace within the sampling system in presence and absence of a biological sample, and on its collection in sorbent tubes to be subsequently analysed by TD-GC/MS.

A small empty Petri dish (Merck Chemicals Ltd, Nottingham, UK) (where the sputum was placed in case of collection of headspace from the sputum samples) was located, uncovered, into a custom vessel (University of Leicester, UK) made of polyether ether ketone (PEEK), a semi-crystalline thermoplastic material with excellent mechanical and

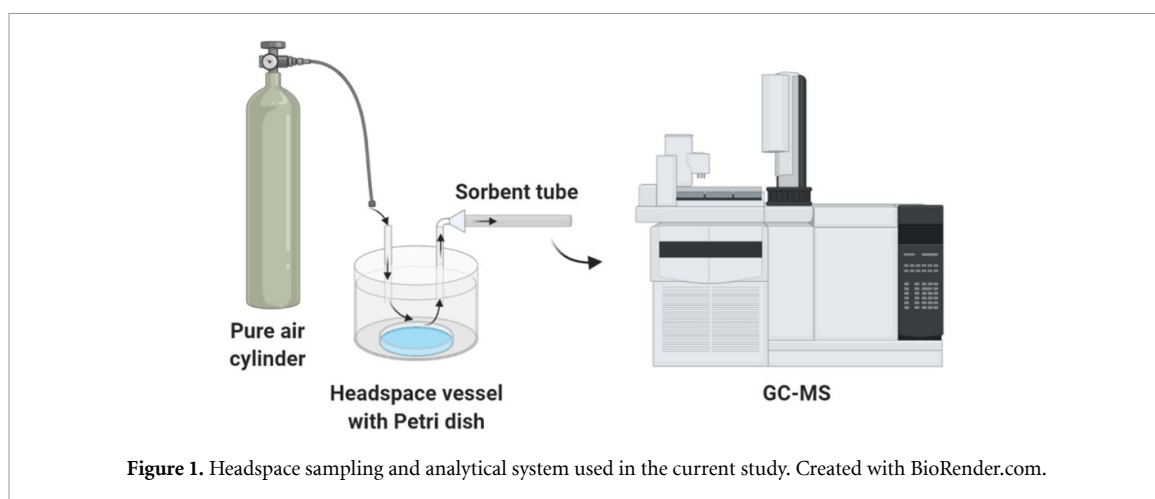
chemical resistance properties. The vessel was sealed with a screw cap fitted with two ports; an inlet port, joined to a pure air cylinder and an outlet port which, during the headspace sampling, was connected to the sorbent tube as reported in figure 1. In order to generate the dynamic headspace, the empty Petri dish contained in the vessel was flushed with a flow of air (BTCA grade, BOC, UK) at a flow rate of 200 ml min⁻¹ for 5 min, sampling 1 l of air onto the sorbent tube. The airflow displaced the headspace above the Petri dish and concentrated VOCs from the headspace onto the sorbent tube (carbograp 1TD/tenax TA 60/40, Markes International, Llantrisant). The air composition was 20.04% oxygen with nitrogen balance (total hydrocarbons ≤0.1 ppm, carbon monoxide ≤1 ppm, carbon dioxide ≤300 ppm, total NO_x ≤0.1 ppm).

2.3. Sorbent tubes preparation

The sorbent tubes were previously weighed and conditioned; the uncapped tubes were weighed before each conditioning procedure to check possible sorbent mass degradation and were conditioned at 330 °C using a nitrogen flow at a pressure of 1.5 bar, for 150 min. Before the TD-GC-MS analysis, in order to ensure that inter-samples variability was maintained at acceptable levels and analytical system artefacts were minimised, an internal standard solution was loaded into the sorbent tube. This was performed by attaching the sorbent tube to an injector and by introducing 0.6 µl of 20 µg ml⁻¹ of internal standard solution in a purified nitrogen flow of 100 ml min⁻¹ for 120 s to ensure an efficient transfer of the standard onto the adsorbent packing. The internal standard solution was a mixture of deuterated toluene (D8), octane (D18) and phenanthrene (D10).

2.4. TD-GC/MS headspace analysis

The background headspace samples were analysed by TD-GC-MS; using a Unity-2 thermal desorption unit (Markes International, Cardiff, UK) linked to a gas chromatograph (Agilent, 7820A) coupled to a single quadrupole mass spectrometer (Agilent, 5977B). The sorbent tubes were placed inside the thermal desorption unit and desorbed at 300 °C for 5 min with a flow rate of 45 ml min⁻¹ splitless. The volatiles were drawn along the heated line onto a 'hydrophobic, general' trap (matching the sorbent in the sample tubes) which was held at -10 °C. The trap was heated at the maximum heating rate to 300 °C for 5 min with a desorption flow rate of 2 ml min⁻¹ splitless and the compounds were introduced to the GC column (Rxi-5ms 60 m × 0.25 mm i.d. × 0.25 µm). The volatiles were separated on the chromatographic column in a Helium flow of 2 ml min⁻¹ with an initial oven temperature of 40 °C and a final oven temperature of 300 °C hold for 5 min (oven temperature ramp of 5 °C min⁻¹). Mass spectrometer was fitted out of an electron ionization source and a single quadrupole as



mass filter. The mass scanning was performed for the range of 40–450 m/z and a frequency of 3 Hz.

2.5. TD-GC-MS data analysis

The TD-GC-MS data analysis was performed in AnalyzerPro (SpectralWorks, Runcorn, UK—version 5.7) for integration, deconvolution and alignment of chromatographic peaks across the samples. The software annotated each feature with its Kovats' retention index (RI), the first and second mass spectrum quantifier ion (m/z) and its retention time, which together constituted a unique compound label that was used to align the identified compounds across both the headspace samples collected at different time-points in a day and those sampled on different days in a week. To the chromatographic features was assigned a

Level 2 annotation in accordance with the Metabolomics Standards Initiative [23] based on mass spectral similarity with the NIST 2011 mass spectral library, using a 50% confidence threshold, and on matching the RI with data reported using the same chromatographic column stationary phase. The chromatographic feature identities were compared to data acquired for chemical standards (Sigma-Aldrich), where available, run on the TD-GC-MS in order to assign to the chromatographic features a Level 1 of identification through comparison of RI and mass spectra, aiming to achieve a reliable identification of the compounds of interest. The analytical standards were run on the GC-MS for 23 out of 37 features identified in at least the 80% of all the samples (as reported in table 1, third column). For the remaining features it was not possible to run analytical standards because they were not available or because some features, in absence of a NIST match, were just identified with their RI and base and secondary most intense ions in their mass spectrum.

For the study of background volatiles intra- and inter-day variations in headspace of blank samples the software generated a data matrix containing the chromatographic features (rows) identified across all

the headspace samples (columns) of both the studies and their respective chromatographic peak area values representing their abundance.

The data analysis of sputum headspace and the respective headspace controls generated a matrix of chromatographic features which was used to consult the most abundant background volatiles selected in the previous study.

2.6. Chemometric analysis

The data matrices generated by AnalyzerPro from the GC-MS data analysis of both the intra- and inter-days background headspace variation studies, and for the sputum headspace study were edited removing VOCs deriving from the internal standard solution spiked into sorbent tubes previous GC-MS analysis. The abundance value of each VOC was normalised for the internal standard toluene D8—the most representative internal standard across all the samples, compared to the others spiked in sorbent tubes before thermal desorption. For the intra- and inter-days background headspace variation studies the data matrix contained a list of 1120 chromatographic features across the all the 35 headspace samples, 20 samples collected in quadruplicate at five different time-points in a day, and 15 collected in triplicate on different days in a week.

A list of features characterising the background signal generated from the headspace sampling system was obtained analysing the original data matrix and identifying those compounds present in at least the 80% of all the headspace samples. The list of 37 features was used to select 13 compounds for the statistical analyses for both the intra-day and inter-days background variations studies. The 13 selected features met the selection criteria of being the most abundant volatiles—with an abundance above or equal to the average value of all the peak areas of the VOCs found in at least the 80% of all the headspace samples—and being ubiquitously present in all the samples collected at each time-point in a day, on each day in a week and in all the replicates.

Table 1. List of 37 chromatographic features detected in at least the 80% of all the background headspace samples.

Chromatographic features ^a	NIST library match ^b	MSI levels ^c	RT ^d	Formula	CAS registry number ^e	HMDB (headspace sample—disease) ^f
693.66_43_58	Acetone	1	3.45	C ₃ H ₆ O	67-64-1	<ul style="list-style-type: none">• Faeces—Crohn’s disease [24]• Breath—Asthma [25]• Saliva—Celiac disease [26]
729.84_78_77	Benzene	1	4.60	C ₆ H ₆	71-43-2	
754.23_74_43	Butanoic acid, methyl ester	1	5.37	C ₅ H ₁₀ O ₂	623-42-7	<ul style="list-style-type: none">• Faeces—Ulcerative colitis [27]
780.61_91_92	Toluene	1	6.21	C ₇ H ₈	108-88-3	<ul style="list-style-type: none">• Breath—Asthma [20]• Faeces—Non-alcoholic fatty liver disease [21]
799.76_56_43	Hexanal	1	6.82	C ₆ H ₁₂ O	66-25-1	<ul style="list-style-type: none">• Faeces—Ulcerative colitis [27]• Cerebrospinal fluid—Hepatic encephalopathy [28]
822.65_207_208	Cyclotrisiloxane, hexamethyl-	2	7.28	C ₆ H ₁₈ O ₃ Si ₃	541-05-9	—
873.15_91_106	p-Xylene	1	8.59	C ₈ H ₁₀	106-42-3	<ul style="list-style-type: none">• Breath—Asthma [29]• Faeces—Ulcerative colitis [27]
893.93_104_103	Styrene	1	9.17	C ₈ H ₈	100-42-5	<ul style="list-style-type: none">• Faeces—Ulcerative colitis [27]
900.76_57_43	Nonane	1	9.35	C ₉ H ₂₀	111-84-2	<ul style="list-style-type: none">• Faeces—IBS [30]• Breath—Asthma [31]
907.59_57_45	na	x	9.54	na	na	—

(Continued)

Table 1. (Continued).

Chromatographic features ^a	NIST library match ^b	MSI levels ^c	RT ^d	Formula	CAS registry number ^e	HMDB (headspace sample—disease) ^f
1049.42_77_106	Benzaldehyde	1	11.15	C ₇ H ₆ O	100-52-7	• Faeces—Ulcerative colitis [30]
1050.46_57_43	Decane, 6-ethyl-2-methyl-	2	11.217	C ₁₃ H ₂₈	62108-21-8	—
1057.40_94_66	Phenol	1	11.65	C ₆ H ₆ O	108-95-2	• Saliva—Celiac disease [26] • Faeces—Crohn's syndrome [24]
1059.72_118_117	α-Methylstyrene	1	11.80	C ₉ H ₁₀	98-83-9	—
1065.37_281_282	Trisiloxane, 1,1,1,5,5,5-hexamethyl-3-[(trimethylsilyl)oxy]-	2	12.16	C ₉ H ₂₇ O ₃ Si ₄	1873-89-8	—
1067.07_57_43	Decane	1	12.27	C ₁₀ H ₂₂	124-18-5	• Breath—Allergic asthma [31]
1068.13_41_43	Octanal	1	12.35	C ₈ H ₁₆ O	124-13-0	• Faeces -Treated celiac disease [32] • Saliva—Celiac disease [26] • Ulcerative colitis [27]
1080.33_68_67	D-Limonene	1	13.23	C ₁₀ H ₁₆	5989-27-5	• Breath—Allergic asthma [33]
1095.65_105_77	Acetophenone	2	14.34	C ₈ H ₈ O	98-86-2	• Breath—Asthma [34]
1103.36_121_43	Benzenemethanol, α, α-dimethyl-	2	14.90	C ₉ H ₁₂ O	617-94-7	—

(Continued)

Table 1. (Continued).

Chromatographic features ^a	NIST library match ^b	MSI levels ^c	RT ^d	Formula	CAS registry number ^e	HMDB (headspace sample—disease) ^f
1108.62_57_43	Undecane	1	15.28	C ₁₁ H ₂₄	1120-21-4	<ul style="list-style-type: none">• Breath—Asthma [29]• Faeces—Clostridium difficile infection [27]
1309.60_57_41	Nonanal	1	15.40	C ₉ H ₁₈ O	124-19-6	<ul style="list-style-type: none">• Breath—Allergic asthma [31]• Saliva—Celiac disease [26]
1309.60_57_71	Nonane, 1-iodo-	2	15.54	C ₉ H ₁₉ I	4282-42-2	—
1309.60_128_127	Naphthalene	2	17.94	C ₁₀ H ₈	91-20-3	<ul style="list-style-type: none">• Breath—Asthma [34]
1309.60_73_267	Cyclopentasiloxane, decamethyl-	2	16.873	C ₁₀ H ₃₀ O ₅ Si ₅	—	—
1310.55_57_43	Tridecane	1	18.21	C ₁₃ H ₂₈	629-50-5	<ul style="list-style-type: none">• Breath—Asthma [29]
1343.43_135_108	Benzothiazole	1	19.09	C ₇ H ₅ NS	95-16-9	—
1351.43_55_141	Cyclohexane, 2-isothiocyanato-	1	19.31	C ₇ H ₁₁ NS	1122-82-3	—
1395.98_71_57	Dodecane, 2,7,10-trimethyl-	2	20.50	C ₁₅ H ₃₂	3891-98-3	—
1415.24_57_43	Tetradecane	1	21.02	C ₁₄ H ₃₀	629-59-4	<ul style="list-style-type: none">• Breath—Allergic asthma [31]
1417.22_142_141	Naphthalene, 2-methyl-	2	21.07	C ₁₁ H ₁₀	91-57-6	—
1514.48_57_71	Pentadecane	1	23.67	C ₁₅ H ₃₂	629-62-9	<ul style="list-style-type: none">• Saliva—Celiac disease [26]

(Continued)

Table 1. (Continued).

Chromatographic features ^a	NIST library match ^b	MSI levels ^c	RT ^d	Formula	CAS registry number ^e	HMDB (headspace sample—disease) ^f
1620.02_73_147	Cycloheptasiloxane, tetradecamethyl-	2	26.18	C ₁₄ H ₄₂ O ₇ Si ₇	107-50-6	—
1666.70_43_57	3,4-Hexanedione, 2,2,5-trimethyl-	2	27.29	C ₁₀ H ₁₈ O ₂	4388-88-9	—
1719.85_57_71	Hexadecane	1	28.56	C ₁₆ H ₃₄	544-76-3	—
1786.83_73_147	Cyclooctasiloxane, hexadecamethyl-	2	30.156	C ₁₆ H ₄₈ O ₈ Si ₈	556-68-3	—
1918.04_178_176	na	x	32.91	na	na	—

^a Feature labels: retention index (RI)_first quantifier ion (m/z)_second quantifier ion (m/z).
^b MSI: Metabolomics Standards Initiative.
^c NIST: National Institute of Standard Technology.
^d RT: retention time.
^e CAS: chemical abstract service.
^f HMDB: Human Metabolome Database search to report the potential biomarker role of some VOCs in different diseases.

Eleven of the thirteen background selected features were identified in sputum headspace samples and their controls (headspace in absence of biological sample) and used to evaluate background contribution in sputum headspace signal of severe asthmatic patients.

2.7. Statistical analysis

To analyse the background headspace intra- and inter-day abundance variation, one way analysis of the variance for the normalised abundance values, summarised as median and interquartile range, was examined by Kruskal–Wallis test for each selected feature and the inter-group comparisons (between time-points or between days) were performed by Dunn's multiple comparisons test. A p value < 0.05 was required to be considered statistically significant.

Z-scores were calculated to standardise compounds abundance values (for the selected group of 13 features) to monitor their variations from the mean. A linear regression plot was fitted to the features z-scores to evaluate the functional relation between abundance of background compounds in headspace and the sampling times. The standard statistical comparator tests and graphical plots with the exclusion of principal component analysis (PCA), were performed using GraphPad Prism 8 (GraphPad Software Inc. La Jolla, CA, USA).

PCA was conducted for the data matrix of 37 features normalised abundance values identified in all the 35 headspace samples to generate a 3D scatter plot to visualize similarities between samples; the first three principal components were plotted in the Cartesian planes (x , y , z). PCA identified the major abundance variations of the selected chromatographic features in background headspace in the main principal components. The PCA was performed using IBM SPSS (Version 26.0. Armonk, NY: IBM Corp.).

The PCA on the abundance values of the 1120 features instead, was automatically generated by AnalyzerPro for both the studies.

In the study to evaluate the background contribution into sputum headspace samples, the normalised peak area values of 11 of the 13 selected volatiles (due to the absence of two of the 13 selected features in sputum headspace samples)—identified in both the headspace of control (blank) samples and headspace of sputum samples—were summarised as mean and standard deviation (SD) and analysed by a multiple t test (two tailed and considered significant at $p < 0.05$) using GraphPad Prism 8 to compare abundance of each background feature between control and sputum headspace. GraphPad Prism 8 was used also to plot the normalised chromatographic peak area average values ratio between sputum headspace and control headspace samples for each of the selected background features.

3. Results

For the intra- and inter-days background headspace variation studies the GC-MS data analysis for all the headspace background samples generated a matrix of chromatographic features that contained the peak area values of 1120 chromatographic features across the headspace samples collected at each time-point (11am to 3pm) in a day, on each day (Monday–Friday) in a week and in all the replicates, resulting in a matrix of 1120 compound rows (\times) 35 columns of samples. As shown in figure 2(a), 37 features were present in at least the 80% of all the 35 background headspaces and were considered in this study as the best representative constituents of the background signal generated from the headspace sampling system. The 80% cut off was arbitrary chosen to select those features stably recurrent in background headspace samples.

The remaining features were present in less than the 80% of the background headspace samples; their sporadic appearance in headspace of blank samples suggested that, although representing part of the background signal, they probably originated from sources which were external to the headspace sampling system. In order to estimate the variability of the background signal generated from the sampling system we referred to the most constant identified features.

The PCA results for the data matrix of 37 features in all the background headspace samples are summarised in the 3D scatter plot in figure 2(b). The total proportion of variance explained by the three principal components was 93.9%. PCA identified that features from the intra-day study clustered around components 1 and 2, whereas, features from the inter-days study clustered around components 2 and 3.

In table 1 the identities of the selected 37 chromatographic features are reported as AnalyzerPro automatic generated labels—consisting of compound Kovats' RI and first and second quantifier ions (m/z)—and as NIST 2011 library match; the features whose identification was confirmed by TD-GC-MS analysis of standards are marked in the third column. The Human Metabolome Database (HMDB) consultation allowed to find references in literature that reported the listed background chromatographic features as potential diseases biomarkers; many of them were identified in breath and saliva, faeces, urine headspace as potential biomarkers for asthma, several chronic gastrointestinal diseases and cancer [22–29].

To study the intra-day and inter-days variations of the background signal, 13 chromatographic features were selected from this table—because they were ubiquitously present in all the headspace samples collected at each time point in a day, on each day in a week and in all the replicates with a peak area value above

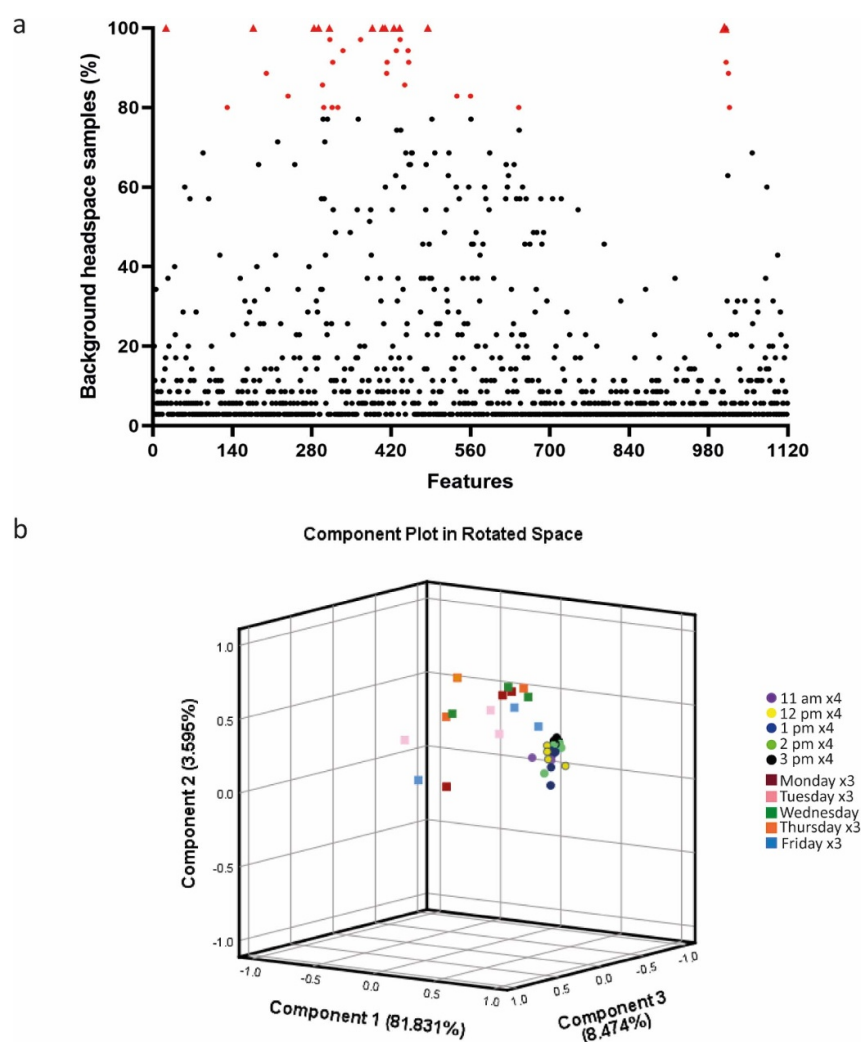


Figure 2. (a) Percentage of the samples in which each of the 1120 features was detected. In red, volatiles present in at least the 80% of the background headspace samples (the red triangles represent the 13 features present in 100% of the samples). (b) 3D scatter dot plot of a principal component analysis (PCA) of a data matrix of 37 selected background volatiles identified in 35 headspace samples. PCA discriminated the 20 background headspaces collected in a single day hourly from 11am to 3pm from the 15 collected on different weekdays from Monday to Friday.

or equal the average of all the values—and used for statistical analysis.

For the study of the background signal intra-day variations, the PCA in figure 3(a), which was performed on the abundance values of the 1120 background features, showed samples grouping according to their collection time in a day; the score plot showed that the samples collected at 11am and 3pm were outliers whereas the other samples were not all loaded on PC1 primarily.

The results for the statistical analysis of background signal intra-day variations at different time-points are shown in the bar plot in figure 3(b); the normalised peak area values of the 13 selected background features were plotted at five different time-points in a day. The analysis of the variance showed a significant difference for the intra-day abundance variations of benzothiazole only, between 11am and 3pm (11am–3pm; $\alpha = 0.0034$). However, the fold change for benzothiazole abundance values between

11am and 3pm was small (2.13). A graphical visualization of the abundance variation of the selected components across the different time-points in a day is shown in figure 3(c); the peak area values reported as z-scores, displayed the compound abundance variations enclosed within ± 1.96 SD.

For the background VOCs inter-day variation study, the PCA in figure 4(a) showed the 15 background headspace samples grouped according to the day of the week in which they were collected and based on the abundance values of all the 1120 features. The results of the statistical analysis for the inter-day variation of background headspace features are reported in figure 4(b); no statistically significant difference in the abundance values of the 13 selected chromatographic features was reported between different week days. The inter-day variations of the abundance values reported as z-scores were enclosed within ± 1.96 range, as showed in figure 4(c).

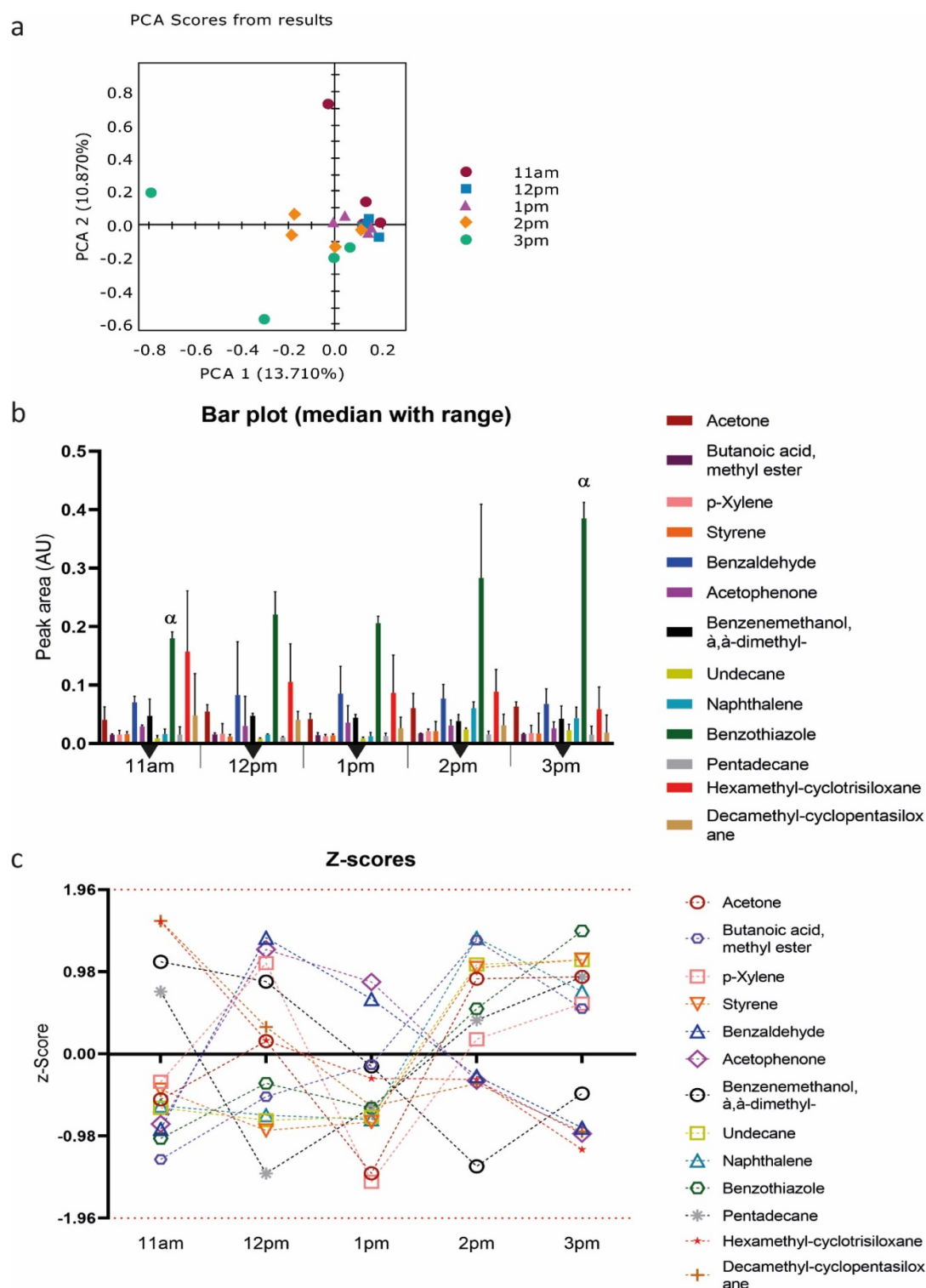
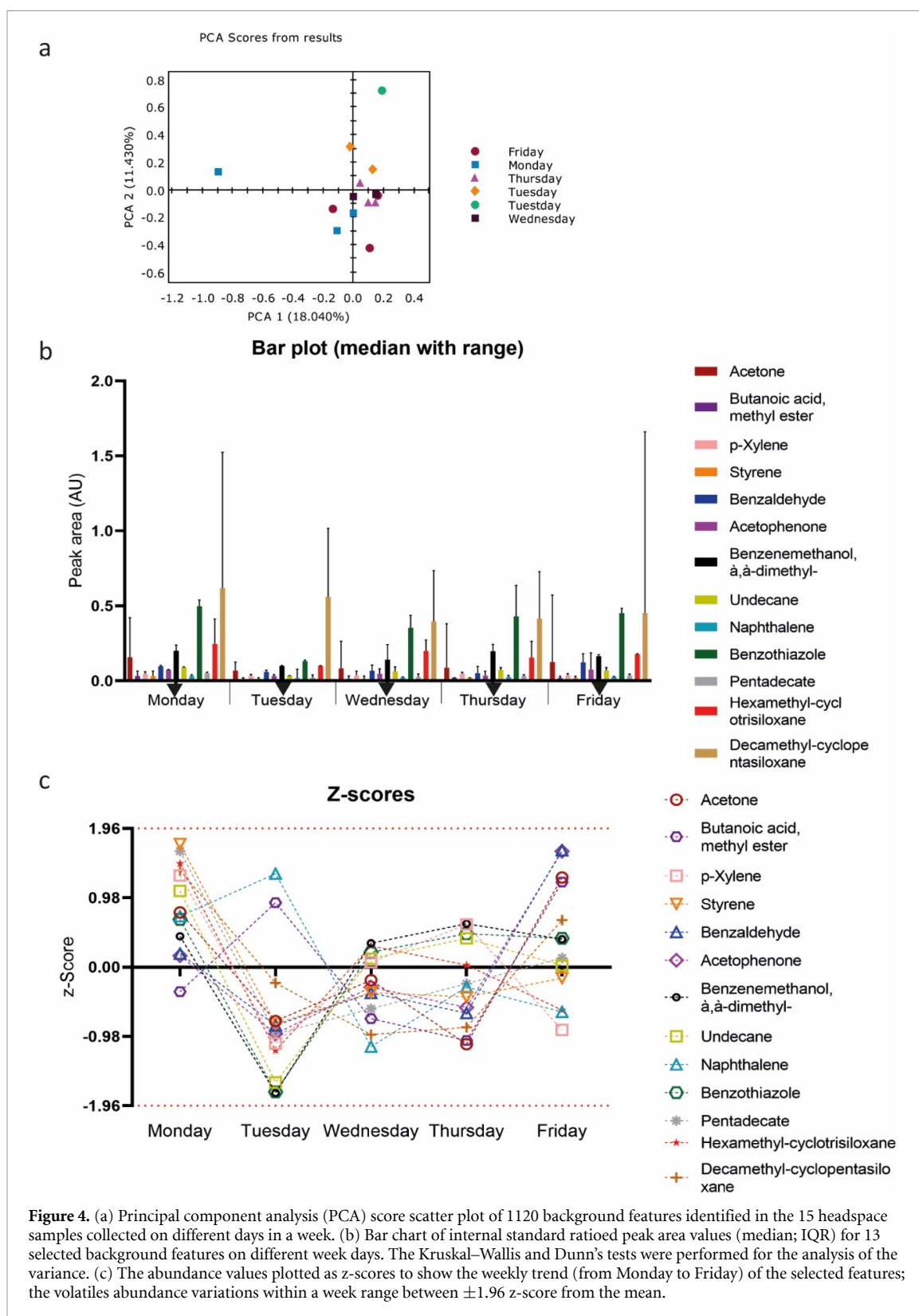


Figure 3. (a) Principal component analysis (PCA) score scatter plot of 1120 background features identified in headspace samples collected four times at five different time points. (b) Bar chart of internal standard ratioed peak area values (median; IQR (interquartile range)) for 13 selected background features at different time-points. Benzothiazole: Kruskal–Wallis $p < 0.05$; Dunn's pairwise comparison test $p < 0.05$ (11am–3pm; $\alpha = 0.0034$). (c) The abundance values plotted as z-scores; the daily trend (from 11am to 3pm) of the abundance variations for the selected volatiles ranges within ± 1.96 z scores from the mean.

To test whether the abundance of the selected background chromatographic features changed in relation to the sampling times (to the time-points in a day and to the days in a week), a linear regression plot

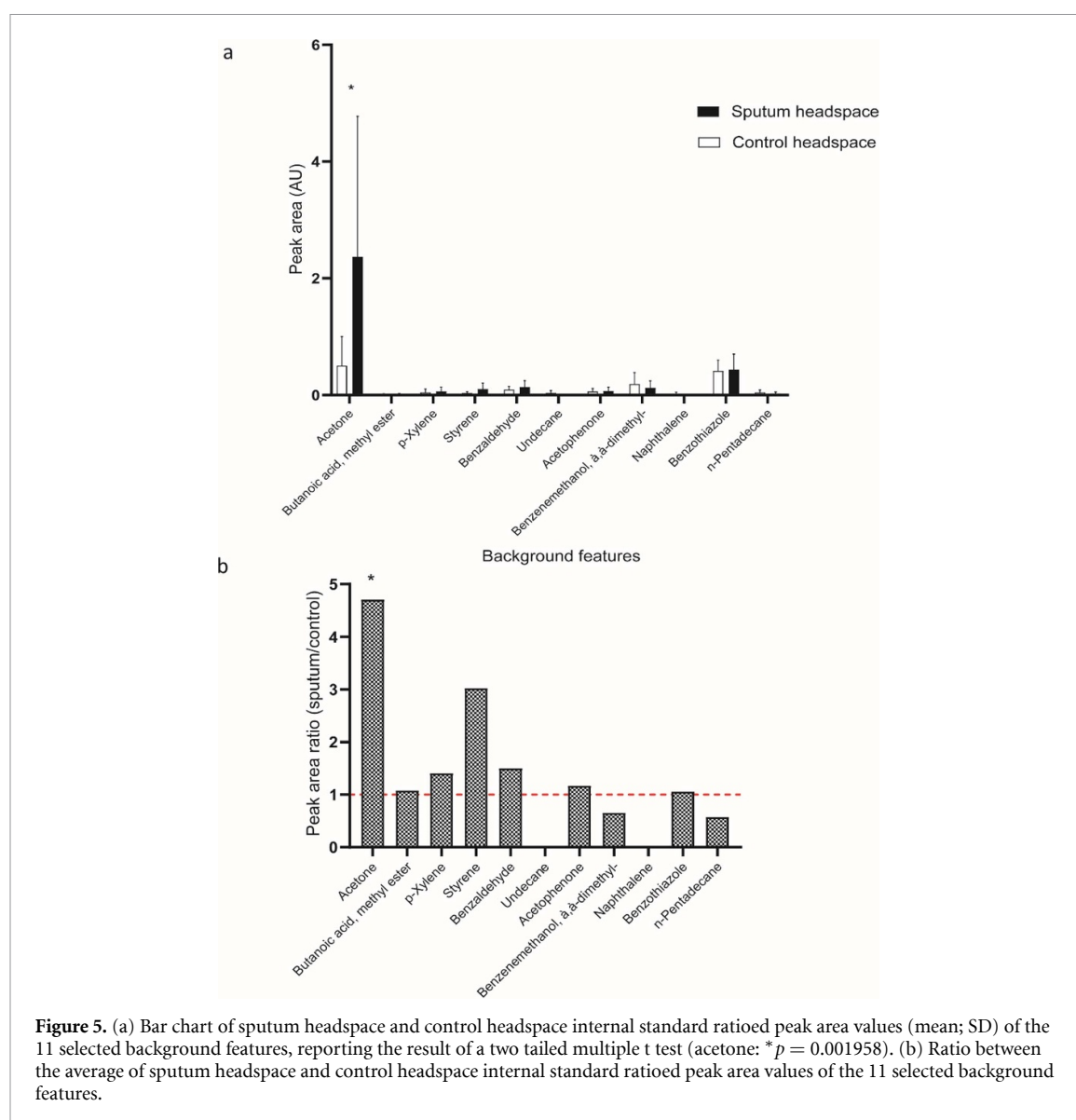
was fitted to both the z-scores plots of both studies; each feature was plotted against each time-point in a day and each day in a week; the only significant linear relation was observed for benzothiazole between



11am and 3pm: linear regression showed a best fit of $y = 0.04446 \times x - 0.3232$, $R^2 = 0.8524$ and p -value = 0.0252.

The analysis of the contribution of 11 of the 13 selected background features into headspace of 36 severe asthma sputum samples and 36 blank controls is shown in figure 5. The normalised peak

area value ratio was calculated for each background feature between sputum headspace and control headspace displaying that the levels of background volatiles increased in sputum headspace compare to control with the exception of benzenemethanol, α , α -dimethyl- and pentadecane whose levels decreased and undecane and naphthalene which were not



detected in sputum headspace. A statistical significant increase of acetone abundance is shown in headspace of biological samples compared to control headspace sample ($p = 0.001958$). Three fold increase of the abundance of styrene was also reported in headspace of sputum samples followed by 1.5 fold increase of benzaldehyde levels; p-xylene, butanoic acid methyl ester-, acetophenone and benzothiazole showed an increase of abundance of <1.5 fold.

4. Discussion

The volatile signal in the headspace of a biological sample contains exogenous volatiles that represent the background signal. This study aimed to characterise the background signal generated within a headspace sampling system developed for collecting headspace from sputum samples and to study its intra- and inter-day variations, and also to evaluate

background signal contribution in headspace of sputum samples collected from severe asthmatic patients. The sampling system consisted of a small PEEK vessel, an empty Petri dish and a supply of pure air. In the study to characterize the background signal and to study its intra- and inter-day variations, background headspace samples were collected in quadruplicate at five time-points in a day and in triplicate on five week days in order to study their respectively intra-day and inter-days abundance variations.

We have made three key observations in this study. Firstly, that the variation of the most abundant features found in the background of headspace capture system did not appear (with the exception of a single feature) to show significant variations within and between days suggesting that routine acquisition of samples in clinical studies may not necessarily need to be standardised to a particular time or day (for the time windows evaluated in this study). This assertion is only likely to be valid for features that are abundant

in the majority of samples as reported here, as we demonstrated clustering features according to time of the day and day of the week when applying PCA to all 1120 quantified features.

Secondly, we identify that many of the most abundant features in the background system are widely reported as disease biomarkers in inflammatory disorders such as asthma and inflammatory bowel disease. Although we noted that the features found in headspace of blank samples originate from exogenous sources, this does not exclude them as putative biomarkers in breath or headspace of biological samples. An exogenous compound present in the inspired air can be also endogenously produced and released in alveolar breath; this may cause the potentially unpicking of endogenously derived components as biomarkers. For example, although in this study acetone represents one of the most abundant background features in headspace of blank samples, it is also commonly produced endogenously in human body and released in breath and in sputum headspace. Acetone abundance variations in breath can be associated to pathological processes such as diabetic ketoacidosis [35] and have also been described as disease biomarkers with asthma [20]. The acetone detected in this study shows higher concentration in headspace of sputum samples compared to its levels in headspace of blank samples, as showed in figure 5.

There are as well exogenous compounds in the inspired air which are degraded and/or excreted in the body and that therefore may be present in low concentration in alveolar breath unless a pathological condition, affecting their metabolism, occurs. For this reason some exogenous compounds—normal constituents of the background signal in headspace and breath samples, resulting from dietary or environmental exposure [36]—may represent potential disease biomarkers. For example, limonene is a compound present in at least the 80% of the headspace samples collected from the sampling system in our study, but although it is an exogenous compound, it is adsorbed by human body and metabolised in the liver and in case of liver dysfunction it is released in breath. Ferrandino *et al* have reported limonene as a putative volatile biomarker of cirrhosis and hepatocellular carcinoma (HCC), showing a correlation between limonene levels in breath and the cirrhosis-induced HCC compared to the healthy controls [37].

Finally we demonstrate that several background VOCs, previously described as asthma biomarkers are present and enriched in the headspace of severe asthmatic sputum samples (e.g. acetone and acetophenone), whereas other putative asthma biomarkers whilst present in sputum headspace were not enriched compared to background samples e.g. Naphthalene (which in this case was undetected in sputum headspace).

The majority of the 1120 identified background features were found in a small percentage of the

total number of background headspace samples (figure 2(a)); therefore a group of 37 features, which were the most frequently identified among all the headspace samples, was selected as representative of the background signal generated from the headspace sampling system and in absence of a biological sample.

The HMDB consultation for the 37 background features in table 1 reported many studies in which some of these background volatiles, in concomitance with other compounds, were identified in headspace of biological samples and in breath, as part of diseases biomarker panels. Background features identified in this study, such as butanoic acid methyl ester, p-xylene, styrene and octanal, were detected in headspace of faeces samples as potential biomarkers of ulcerative colitis [27]; phenol, octanal, nonanal, pentadecane, and benzene were identified in saliva headspace samples as potential biomarkers of celiac disease [26], while tetradecane, decane and nonanal, were reported in a panel of potential breath biomarkers of allergic asthma [31]. The background features identified in headspace of blank samples may originate selectively from the different components of the headspace sampling system.

The vessel used for the headspace collection is made of PEEK, which is a high-performance semi-crystalline engineering thermoplastic. The PEEK's aromatic backbone comprising the bulk of its monomer unit—which attributes to the PEEK excellent thermal properties—could be, for instance, the source of aromatics components; moreover PEEK can release propene (not observed in this study) and p-xylene [38] which has been widely detected in the headspace of blank samples.

The flexible tube that links the vessel to the gas cylinder and the sorbent tubes holders are made of polyfluor tetraethylene, the most common man-made chemical belonging to the Teflon® group, which could be responsible for the presence of fluorinated components in the headspace samples such as the 1-1-difluoroethane but also of alcohols like the 2-ethyl-1-hexanol [38].

However, both PEEK and Teflon®, commonly used to make flexible tubing as part of the headspace collection system, are reported in literature as the most stable materials when comparing their volatilome to the one produced by other substances such as the polyvinyl chloride, nylon, silicone and polyethylene [39].

The polystyrene Petri dishes used in this study can release styrene, while the active carbon filter contained in sorbent tubes can be the source of hydrocarbons [40], which is the reason why the sorbent tubes are conditioned in nitrogen previously their use; besides, the oxidation of the TENAX TA sorbent can release benzaldehyde [41].

All these observations stress the importance of a good knowledge of the background signal generated

from the collecting system in headspace samples and highlight the potential impact of the background signal and its abundance variations on hypothetical panels of potential disease biomarkers.

In this study it was demonstrated, that although the 13 selected background features showed a wide intra-day as well as inter-day abundance variations, these changes were not statistically significant and therefore it seems unlikely that the selected background features can have a relevant impact on the composition of a potential panel of biomarkers developed from headspace samples. Only benzothiazole showed a significant intra-day linear abundance increase between 11am and 3pm. However, the fold change in benzothiazole normalised peak area value between 11am and 3pm was 2.13, indicating just a small numerical increase in the abundance value across the day and the variations were enclosed within ± 1.96 SD from the mean. These results suggested that it is unlikely that the variation of benzothiazole levels could have a relevant impact on the development of a potential biomarker signature in headspace of biological samples, in addition it has to be considered that it is unusual, in the process of biomarker signature development and validation, to rely just on a single biomarker for discrimination [42].

The inter-days abundance variations of the identified features in background headspace samples did not showed any relevant results, and the level of benzothiazole in this experiment did not resulted significantly different between different weekdays (figure 4(b)). Moreover, there was no linear relationship between the z-score of abundance values of the selected chromatographic features and the different days of the week in which the background headspace samples were collected.

The study has a number of limitations. Firstly, the intra-day study only considered a narrow series of time intervals and it is known that VOCs demonstrate diurnal variation [43]. However, the time intervals and weekday sampling were chosen as it coincides with when routine clinical samples for headspace analyses would normally be collected. Secondly, we did not localise the background signals to individual components of the sampling system. Future studies will need to evaluate individual system components to try to minimise further background VOCs.

The subsequent study has showed that 11 features reported in the background, and originated from the headspace sampling system, were present and enriched in headspace of sputum samples acquired from patients with severe asthma compared with headspace of control samples (headspace collected in absence of biological samples) (figure 5). The increased levels of acetone in headspace of sputum samples were expected considering that acetone is also an endogenous compound, and one of the most

abundant volatile identified in breath samples of both asthmatics and non-asthmatics [25].

Styrene is the second background feature that showed a 3 fold increase in headspace of sputum samples even though the difference in abundance levels between sputum and control headspace samples was not reported as statistically significant. Styrene is a volatile monomer used in the production of polymers, copolymers and reinforced plastics and it has been reported as cause of occupational asthma; workers exposed to a polyester resin containing styrene showed a worsening of upper airway and asthma symptoms [44]. These results suggested that styrene is released not only from the headspace sampling system, but it is probably also present in ambient air and internalised in human body and released in biological samples. The study revealed also a decrease in abundance levels of benzenemethanol, α , α -dimethyl- and pentadecane in sputum samples compared to controls but the difference between the two groups were not statistically relevant, while the absence of undecane and naphthalene was reported in headspace of sputum samples even though serum levels of naphthalene have been reported to be significantly higher among asthmatics compared to non-asthmatics [45].

This research was exclusively focused on the evaluation of the background signal contribution in headspace of sputum samples generated from a specific sampling system and from a single disease cohort (severe asthmatics patients); it would be interesting in future to evaluate the background signal for other types of collection system for headspace and the contribution in headspace of sputum samples from other disease cohorts and healthy volunteers to better understand the specificity of putative biomarkers.

We acknowledge the potential of a selection bias given high proportion of males to females for asthma cohorts. However, it should be noted that the acquisition of spontaneous sputum (without hypertonic saline induction) is conditional on patients generating sputum plugs when they are stable (i.e. not exacerbating), this factor would not be expected to follow population level asthma gender proportions as it is dependent on patients having a bronchitic phenotype of asthma. In additional a significant proportion of our patients were on anti-IL5 biologics (14/30) and it was previously reported that there is a male predominance of patients with late onset eosinophilic asthma [46], who would usually have corticosteroid refractory disease requiring biologic therapies. Finally, gender does not appear to affect the overall profile of exhaled VOCs [47].

Moreover, all of our observations are only likely to be valid for features that are abundant in the majority of samples, therefore further research is needed to study the role and the influence of the highly abundant but less frequent (not present in all the headspace samples) background VOCs.

5. Conclusions

We report the volatome of a headspace sampling system for *in vitro* studies and its intra/inter-day variation. We have identified limited variation of abundant chromatographic features and clustering of less abundant features according to sampling time and day. We have also demonstrated that many of the volatiles identified in the background system have been reported as disease biomarkers. The background signal generated from the headspace sampling system was identified in headspace of sputum samples of severe asthmatics patients and its contribution to the biological signal has been reported. Further studies are required to understand how our observations can be used to study the volatiles headspace of *in vitro* samples acquired in clinical biomarker profiling studies.

Acknowledgments

This research was funded by the Medical Research Council (MRC), Engineering and Physical Sciences Research Council (EPSRC) Stratified Medicine Grant for Molecular Pathology Nodes (Grant No. MR/N005880/1), Midlands Asthma and Allergy Research Association (MAARA) and British Lung Foundation (Grant No. BLFPHD17-1). The work was carried out at the University Hospitals of Leicester NHS Trust, University of Leicester and Loughborough University, supported by the NIHR Leicester Biomedical Research Centre and the NIHR Leicester Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The authors would like to acknowledge the invaluable efforts of the research nurses responsible for the in-clinic sample collection as well as the input from the wider East Midlands Breathomics Pathology Node consortium (members list can be found at: <https://ember.le.ac.uk/web>).

ORCID iDs

Rosa Peltrini  <https://orcid.org/0000-0003-2671-8996>

Dahlia Salman  <https://orcid.org/0000-0002-5354-2407>

C L Paul Thomas  <https://orcid.org/0000-0003-4631-6417>

References

- [1] Abaffy T, Duncan R, Riemer D D, Tietje O, Elgart G, Milikowski C and DeFazio R A 2010 Differential volatile signatures from skin, naevi and melanoma: a novel approach to detect a pathological process *PloS One* **5** e13813–e
- [2] Monedeiro F, Milanowski M, Ratiu I-A, Zmysłowski H, Ligor T and Buszewski B 2019 VOC profiles of saliva in assessment of halitosis and submandibular abscesses using HS-SPME-GC/MS technique *Molecules* **24** 2977
- [3] Schmidt K and Podmore I 2015 Current challenges in volatile organic compounds analysis as potential biomarkers of cancer *J. Biomark* **2015** 981458
- [4] Poole C F 2012 *Headspace-Gas Chromatography* ed M J Sithersingh and N H Snow (Waltham, MA: Elsevier)
- [5] Serasanambati M, Broza Y Y, Marmur A and Haick H 2019 Profiling single cancer cells with volatolomics approach *iScience* **11** 178–88
- [6] Schleich F N, Dallinga J W, Henket M, Wouters E F, Louis R and Van Schooten F J 2016 Volatile organic compounds discriminate between eosinophilic and neutrophilic inflammation *in vitro* *J. Breath Res.* **10** 016006
- [7] Jiang J, Malavia N, Suresh V and George S C 2009 Nitric oxide gas phase release in human small airway epithelial cells *Respir. Res.* **10** 3
- [8] Goeminne P C, Vandendriessche T, Van Eldere J, Nicolai B M, Hertog M L A T M and Dupont L J 2012 Detection of *Pseudomonas aeruginosa* in sputum headspace through volatile organic compound analysis *Respir. Res.* **13** 87
- [9] Timm C M, Lloyd E P, Egan A, Mariner R and Karig D 2018 Direct growth of bacteria in headspace vials allows for screening of volatiles by gas chromatography mass spectrometry *Front. Microbiol.* **9** 491
- [10] Töreyn Z N, Ghosh M, Göksel Ö, Göksel T and Godderis L 2020 Exhaled breath analysis in diagnosis of malignant pleural mesothelioma: systematic review *Int. J. Environ. Res. Public Health* **17** 1110
- [11] 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
- [12] Dhillon K K and Gupta S 2020 *StatPearls* (Treasure Island, FL: StatPearls Publishing)
- [13] Spaněl P and Smith D 2011 Progress in SIFT-MS: breath analysis and other applications *Mass Spectrom. Rev.* **30** 236–67
- [14] Pysanenko A, Spaněl P and Smith D 2008 A study of sulfur-containing compounds in mouth- and nose-exhaled breath and in the oral cavity using selected ion flow tube mass spectrometry *J. Breath Res.* **2** 046004
- [15] Pleil J D, Stiegel M A and Risby T H 2013 Clinical breath analysis: discriminating between human endogenous compounds and exogenous (environmental) chemical confounders *J. Breath Res.* **7** 017107
- [16] Sobus J R et al 2009 Investigation of PAH biomarkers in the urine of workers exposed to hot asphalt *Ann. Occup. Hyg.* **53** 551–60
- [17] Laškova T, Zabukas V and Vaitiekūnas P 2007 Influence of meteorological conditions on volatile organic compound spread in the atmospheric boundary layer *J. Environ. Eng. Landscape Manage.* **15** 135–43
- [18] Boshier P R, Cushnir J R, Priest O H, Marczin N and Hanna G B 2010 Variation in the levels of volatile trace gases within three hospital environments: implications for clinical breath testing *J. Breath Res.* **4** 031001
- [19] Bessonneau V, Mosqueron L, Berrubé A, Muckensturm G, Buffet-Bataillon S, Gangneux J-P and Thomas O 2013 VOC contamination in hospital, from stationary sampling of a large panel of compounds, in view of healthcare workers and patients exposure assessment *PloS One* **8** e55535
- [20] Dragonieri S et al 2007 An electronic nose in the discrimination of patients with asthma and controls *J. Allergy Clin. Immunol.* **120** 856–62
- [21] Raman M et al 2013 Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease *Clin. Gastroenterol. Hepatol.* **11** 868–75.e3
- [22] Reddel H K et al 2019 GINA 2019: a fundamental change in asthma management *Eur. Respir. J.* **53** 1901046

- [23] Salek R M, Steinbeck C, Viant M R, Goodacre R and Dunn W B 2013 The role of reporting standards for metabolite annotation and identification in metabolomic studies *Gigascience* **2** 13
- [24] Walton C *et al* 2013 Analysis of volatile organic compounds of bacterial origin in chronic gastrointestinal diseases *Inflamm. Bowel Dis.* **19** 2069–78
- [25] Smolinska A, Klaassen E M, Dallinga J W, van de Kant K D, Jobsis Q, Moonen E J, van Schayck O C P, Dompeling E and van Schooten F J 2014 Profiling of volatile organic compounds in exhaled breath as a strategy to find early predictive signatures of asthma in children *PloS One* **9** e95668
- [26] Francavilla R *et al* 2014 Salivary microbiota and metabolome associated with celiac disease *Appl. Environ. Microbiol.* **80** 3416–25
- [27] Garner C E, Smith S, de Lacy Costello B, White P, Spencer R, Probert C S and Ratcliffe N M 2007 Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease *FASEB J.* **21** 1675–88
- [28] Goldberg E M, Blendis L M and Sandler S 1981 A gas chromatographic–mass spectrometric study of profiles of volatile metabolites in hepatic encephalopathy *J. Chromatogr.* **226** 291–9
- [29] Dallinga J W *et al* 2010 Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children *Clin. Exp. Allergy* **40** 68–76
- [30] Ahmed I, Greenwood R, Costello B, Ratcliffe N and Probert C S 2016 Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease *Aliment. Pharmacol. Ther.* **43** 596–611
- [31] Caldeira M, Perestrelo R, Barros A S, Bilelo M J, Morêto A, Câmara J S and Rocha S M 2012 Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography *J. Chromatogr. A* **1254** 87–97
- [32] Di Cagno R *et al* 2011 Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization *BMC Microbiol.* **11** 219
- [33] Gahleitner F, Guallar-Hoyas C, Beardsmore C S, Pandya H C and Thomas C P 2013 Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath *Bioanalysis* **5** 2239–47
- [34] van de Kant K D *et al* 2013 Exhaled breath profiling in diagnosing wheezy preschool children *Eur. Respir. J.* **41** 183–8
- [35] Saasa V, Beukes M, Lemmer Y and Mwakikunga B 2019 Blood ketone bodies and breath acetone analysis and their correlations in type 2 diabetes mellitus *Diagnostics* **9** 224
- [36] Gaude E, Nakhleh M K, Patassini S, Boschmans J, Allsworth M, Boyle B and van der Schee M P 2019 Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes *J. Breath Res.* **13** 032001
- [37] Ferrandino G *et al* 2020 Breath biopsy assessment of liver disease using an exogenous volatile organic compound–toward improved detection of liver impairment *Clin. Transl. Gastroenterol.* **11** e00239
- [38] Hayes H C, Benton D J and Khan N Impact of sampling media on soil gas measurements A&WMA ‘Vapor Intrusion—The Next Great Environmental Challenge—An Update’ (13–15 September 2006)
- [39] Mussato B V S and Tsurikova L 2009 *Assessment of Volatile Organic Compounds from Tubing Used for Soil Vapour Sample Collection* (Richmond, BC: CARO Analytical Services)
- [40] Veenaas C, Ripszam M and Haglund P 2020 Analysis of volatile organic compounds in indoor environments using thermal desorption with comprehensive two-dimensional gas chromatography and high-resolution time-of-flight mass spectrometry *J. Sep. Sci.* **43** 1489–98
- [41] Klenø J G, Wolkoff P, Clausen P A, Wilkins C K and Pedersen T 2002 Degradation of the adsorbent Tenax TA by nitrogen oxides, ozone, hydrogen peroxide, OH radical, and limonene oxidation products *Environ. Sci. Technol.* **36** 4121–6
- [42] Seijo L M *et al* 2019 Biomarkers in lung cancer screening: achievements, promises, and challenges *J. Thorac. Oncol.* **14** 343–57
- [43] Brinkman P *et al* 2019 Exhaled volatile organic compounds as markers for medication use in asthma *Eur. Respir. J.* **55** 1900544
- [44] Fernandez-Nieto M, Quirce S, Fraj J, Del Pozo V, Sastre B, Seoane C, Lahoz C and Sastr J 2006 Occupational asthma caused by styrene in a autobody shop worker *J. Allergy Clin. Immunol.* **117** 248–50
- [45] Sudakin D L, Stone D L and Power L 2011 Naphthalene mothballs: emerging and recurring issues and their relevance to environmental health *Curr. Top. Toxicol.* **7** 13–19
- [46] Haldar P, Pavord I D, Shaw D E, Berry M A, Thomas M, Brightling C E, Wardlaw A J and Green R H 2008 Cluster analysis and clinical asthma phenotypes *Am. J. Respir. Crit. Care Med.* **178** 218–24
- [47] 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