The compositional space of exhaled breath condensate and its link to the human breath volatilome

2015 J. Breath Res. 9 027105

(http://iopscience.iop.org/1752-7163/9/2/027105)

Journal of Breath Research

PAPER

The compositional space of exhaled breath condensate and its link to the human breath volatilome

Franco Moritz[1], Monika Janicka[2], Agata Zygler[2], Sara Forcisi[1], Agata Kot-Wasik[2], Jacek Kot[4], István Gebefügi[1], Jacek Namiesnik[2] and Philippe Schmitt-Kopplin[1,3]

[1] Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Department of Environmental Sciences, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany
[2] Gdansk University of Technology, Faculty of Chemistry, Department of Analytical Chemistry, Narutowicza 11/12, 80-233 Gdansk, Poland
[3] Technische Universität München, Chair of Analytical Food Chemistry, Alte Akademie 10, 85354 Freising-Weihenstephan, Germany
[4] Medical University of Gdansk, Institute of Maritime and Tropical Medicine, National Center for Hyperbaric Medicine, Powstania Styczniowego Str. 9B, 81-519 Gdynia, Poland

E-mail: franco.moritz@helmholtz-muenchen.de **and** schmitt-kopplin@helmholtz-muenchen.de

RECEIVED
16 December 2014

REVISED
12 February 2015

ACCEPTED FOR PUBLICATION
23 February 2015

PUBLISHED
6 May 2015

Keywords: exhaled breath condensate, non-targeted metabolomics, volatile organic compounds

S Online supplementary data available from stacks.iop.org/JBR/9/027105/mmedia

Abstract

Breath analysis is commonly understood to target gaseous or volatile organic compounds (VOCs) for the characterization of different pathologies. Targeted analysis is most effective if a working hypothesis can be based on a plethora of data. The recently published volatilome builds an optimal basis for organizing powerful target sets. However, the origin and pathways of biosynthesis of many VOCs are not known, which complicates the formulation of useful hypotheses. To find the missing link between VOCs and their origin, it is necessary to analyze their precursor fluids themselves. In order to provide condensation nuclei for the generation of future hypotheses, we provide the compositional space over 23 samples of the unperturbed human exhaled breath condensate (EBC) metabolome. We propose a way to connect the compositional spaces of both VOCs and EBC so as to gain insight into the most probable form of VOC precursors. In a way analogous to tandem MS it is possible to create a mass difference network over compositional data by linking compositions with mass differences that are designed to mimic biochemical reactions. We propose to use mass difference enrichment analysis (MDEA) in order to mine probable relations between VOCs and their precursor fluids. We have found 2691 EBC compositions and linked them to 235 breath VOC compositions that correspond to 848 individual compounds. We found that VOCs are likely to be found as hexose conjugates or as amino acid conjugates with Glutamine or Asparagine playing a major role. Furthermore, we found that dicarboxylic acid mass differences may be more indicative for oxidative stress than oxygenation-hydrogenation sequences.

1. Introduction

Breath analysis majorly focuses on the sampling, detection and quantification of volatile organic compounds (VOCs) for the purpose of modeling (and diagnosing) diseases or metabolic states [1]. In order to understand the mechanisms behind the relation of VOC profiles or patterns with some phenotype, it is necessary to localize their origin (lung, liver, muscle, heart, adipous tissue, etc) and it is necessary to identify their precursors [2]. Yet, once a precursor is identified, the demarcation of its metabolic sphere of influence is often still a matter of further investigations. In fact, it is the aim of metabolomics on bio-fluids to address such questions [3], and it is to be noted that the metabolic role of a vast majority of compounds detected in non-targeted metabolomics approaches is not known and that there is poor consistency in the definition of metabolic pathways [4]. Many identified precursors and pathways are either related to oxidative stress, or their precursors are known because they are drugs and can be toxicologically traced/screened (e.g. valproic acid and 3-heptanone) [5]. As recently described in de Lacy Costello *et al* [2] the origin of many VOCs is yet to be unveiled. Already given the large amount of VOCs that are known in healthy individuals, it is entirely justified to address this ordeal as a 'gargantuan task'.

In order to identify the precursors of VOCs it is necessary to analyze the sample matrices from which they are presumed to emanate with technologies that are adapted to the liquid phase. Pertaining to breath analysis, relevant precursor-containing bio-fluids are blood plasma and exhaled breath condensate (EBC). Seemingly, VOCs are often hypothesized to be produced in some tissue so as to migrate to breath via the blood stream [2]. The VOCs are hypothesized to exist and to be dissolved in their final form until they finally end up in breath. However, de Lacy Costello *et al* have pointed out that the blood plasma volatilome differs markedly from the breath volatilome. Bearing this in mind, it is to be asked to what extent VOCs are produced *in situ* (in the lung) or *ex situ* (somewhere else). Assuming the *in situ* scenario to be of relevance, it is EBC that needs to be analyzed for the search of immediate VOC precursors, as EBC itself originates from the airway lining fluid [6]. EBC is sampled by condensing exhaled breath in a cooling chamber [7]. In this process condensed H_2O represents the major partition of the sample matrix. Non-volatile compounds (precursors) are caught in EBC because exhaled aerosols, which carry metabolites, proteins [8] and even DNA [9], are dissolved in the condensed water.

Chemical analyses of EBC have been performed for two decades, however not many more than 100 different non-volatile compounds (potential precursors) have been identified and quantified. There are two reasons for this lack of knowledge regarding the chemical constituents of EBC: (a) the water content of EBC varies dramatically, which compromises quantification, and (b) targeted research was mostly focused on inflammation markers whose relevance were reported previously for other sample matrices (e.g. isoprostanes [10]). A wider scope of analytes can be covered by metabolomics [11], however, the majority of metabolomics studies on EBC have been performed using NMR (PubMed search: (Exhaled breath condensate) AND metabolomics). Out of 21 PubMed hits, three papers used mass spectrometric approaches [12–14]; one paper used LC-UV/VIS spectrometry [15]; 11 NMR studies were published [16–26] and seven reviews have been found [16, 21, 27–31]. The numbers do not add up to 21 as NMR was the scope of some reviews. As most studies were performed using NMR, metabolome coverage is limited by the sensitivity of this method.

What delimitates the discovery of new VOC precursors is essentially the limited knowledge about the diversity of EBC compounds, a lack of knowledge about its chemical space and to begin with a lack of knowledge about its compositional space. The compositional space of EBC would be a list of elemental compositions/sum formulas that have been observed in EBC. The chemical space would then describe the identity of each composition. Such knowledge could then be analyzed for precursors in a holistic way; a way that makes the identification of all precursors seemingly less gigantic.

Compositional space analysis is routinely being performed in the field of natural organic matter (NOM) analysis, where complex mixtures are characterized by thousands of elemental compositions that can be mined by using ultra high accuracy and resolution mass spectrometry (UHR-MS) [32, 33]. Instruments such as the Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR/MS) at very high magnetic fields (12 T) are capable of resolving thousands of m/z peaks at once while delivering a mass accuracy (<100 ppb) that enables sum formula assignment without the immediate need of isotopic pattern validation.

Mass accuracy as well enables the direct comparison of compositions from different sample sets by means of e.g. mass difference enrichment analysis (MDEA) [34]. We investigate the compositional space of EBC and compare it to that of the human breath volatilome. We propose the analysis of reaction-mass differences over annotated compositions as a method for VOC-precursor research.

2. Methods

2.1. Chemicals

Methanol (LC/MS grade) and water (LC/MS grade) were purchased from Fluka Analytical (Sigma-Aldrich).

2.2. EBC sampling

EBC was sampled from 24 volunteers using the TurboDeccs System from Medivac (Parma, Italy) as described in Janicka 2012 [10]. Samples were then stored in 1.5 ml Eppendorf tubes (Hamburg, Germany) at −80 °C.

The sample population of volunteers was composed of 12 males and 12 females. The average age was 29 years (standard deviation of 8.6 years). The average body mass index was 23 (standard deviation: 3.7). Except for one volunteer suffering from asthma, all volunteers were healthy.

Sample labels were anonymized; as one sample could not be analyzed, the report at hand encompasses 23 samples. Approval was received from the ethical board of the Medical University of Gdansk.

2.3. Sample preparation

Samples were thawed on ice and one part of EBC was diluted in nine parts of MeOH. Blanks for assessment of carryover were prepared using one part of H_2O and nine parts MeOH.

2.4. Data generation

2.4.1. FT-ICR/MS measurement

Ultra-high accuracy and resolution mass spectra were acquired using a Bruker (Bruker Daltonik GmbH, Bremen, Germany) solariX Fourier transform ion cyclotron resonance mass spectrometer equipped with a 12 T superconducting magnet and an APOLLO II electrospray ion-source in positive ionization mode. Prior to sample acquisition external calibration was

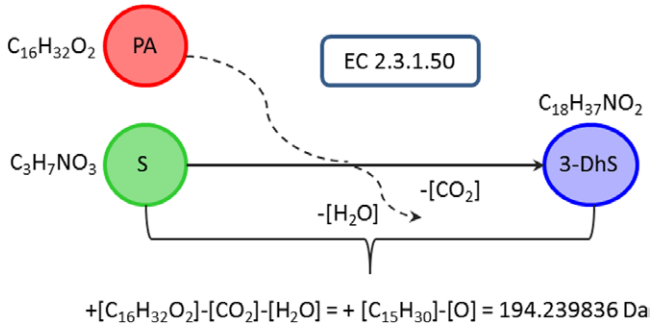

Figure 1. Scheme of the connection of two nodes by an edge in a MDiN. The initial step of sphingosine biosynthesis, catalyzed by the enzyme EC 2.3.1.50 is shown. PA: palmitic acid; S: l-serine; 3-DhS: 3-dehydrosphinganine. Circles are nodes, the full arrow is the edge and the dashed arrow implies the insertion of a theoretical mass difference under decarboxylative condensation. The compositions of S and 3-DhS were annotated in a mass spectrum and their mass difference equals to decarboxylative condensation with PA.

performed on a 1 mg l^{-1} arginine-MeOH solution. Calibration was accepted once a linear calibration curve could be approximated for at least four signals with each individual error being smaller than 100 ppb. Samples were infused into the electrospray ionization (ESI) source in a randomized order with MeOH-H_2O blanks after any five samples. Infusion was performed at a flow rate of 120 µl h^{-1} with a Gilson autosampler system (Gilson, Inc., Meddleton, WI, USA). Spectra were acquired with a time domain of 2 megawords in a mass range of 148 to 1100 *m*/*z* with a resolving power of 400 000 at *m*/*z* 400. Ion accumulation time was set to 1.0 s, time of flight was set to 1 ms. Spectra were accumulated over 405 scans. Capillary voltage and spray shield voltage were set to 4500 V and −500 V, respectively. Drying gas flow rate was set to 4 l min^{-1}. Drying gas temperature was set to 200 °C. Nebulizer pressure was set to 1.1 bar. Raw spectra were processed with Data Analysis Version 4.1 (Bruker Daltonik GmbH, Bremen, Germany). Spectra were internally calibrated with an in house collected list of commonly observed compositions and their respective theoretical *m*/*z* values ($n > 200$). The central error over *m*/*z* distribution was centered around zero ppm. Overall error standard deviation was commonly found to be <100 ppb. Calibrated mass spectra were exported as ASCI files with a signal to noise ratio of 4. Exported peak lists were aligned using an in house written program [35]. The peak intensities of features, which deviated by <1 ppm, were concatenated into an intensity vector and the average exact *m*/*z* was reported for each vector.

2.4.2. The human volatilome

The entire human volatilome as listed in [2] was downloaded and converted to .xls files. The sum formulas of all 1770 non-halogenic VOCs were retrieved manually. 848 breath-VOCs were extracted. In order to acquire the same data basis that is given by FT-ICR/MS, we merged VOCs of equal sum formulas into single variables. This process yielded 235 VOC compositions.

2.5. Data evaluation

2.5.1. Mass difference network reconstruction and annotation of mass spectra

Mass difference networks (MDiNs) were reconstructed as initially proposed by Breitling *et al* [36] and spectra were annotated as in Tziotis *et al* [37]. Reaction equivalent mass differences (REMDs) were based on the functional set as presented in [37]. Further REMDs encompassed exchanges of small functionalities, fatty acids (C2–C16), dicarboxylic acids (from C2–C10), proteinogenic amino acids and their corresponding keto-acids, phenylpropanoids, hexoses, glucuronic acid, glutathione and prenylations (C5, C10, C15, C20). If the mass difference of two *m*/*z* values was found to deviate from an REMD by <200 ppb, they were connected by an edge. An MDiN is reconstructed after all possible *m*/*z* pairs were screened for all REMDs. The general scheme of a MDiN is shown in figure 1.

Sum formula annotation was initiated by assigning isotopically validated sum formulas to the major *m*/*z* peaks. Subsequently, nodes that were connected to a node with known sum formula could be annotated by combining the known sum formula with the information held by the edge. During annotation, all nodes and sum formulas were constantly checked for self-consistency. False assignments were set back to zero and given the chance to yield an agreeable annotation by using an alternative network path. Annotation was finalized once the network annotations reached a local optimum. MDiNs for MDEA were generated using the same REMDs over the theoretical neutral masses of the found composition. The edge error was then set to be zero.

2.5.2. Mass difference enrichment analysis

MDEA tests whether an REMD is overrepresented in its incidence to a group of nodes (a sample population) relative to its overall occurrence in an MDiN. Therefore, it is necessary to use an REMD's occurrence throughout the entire MDiN to compute an expected frequency μ and its expected standard deviation σ for the sample population.

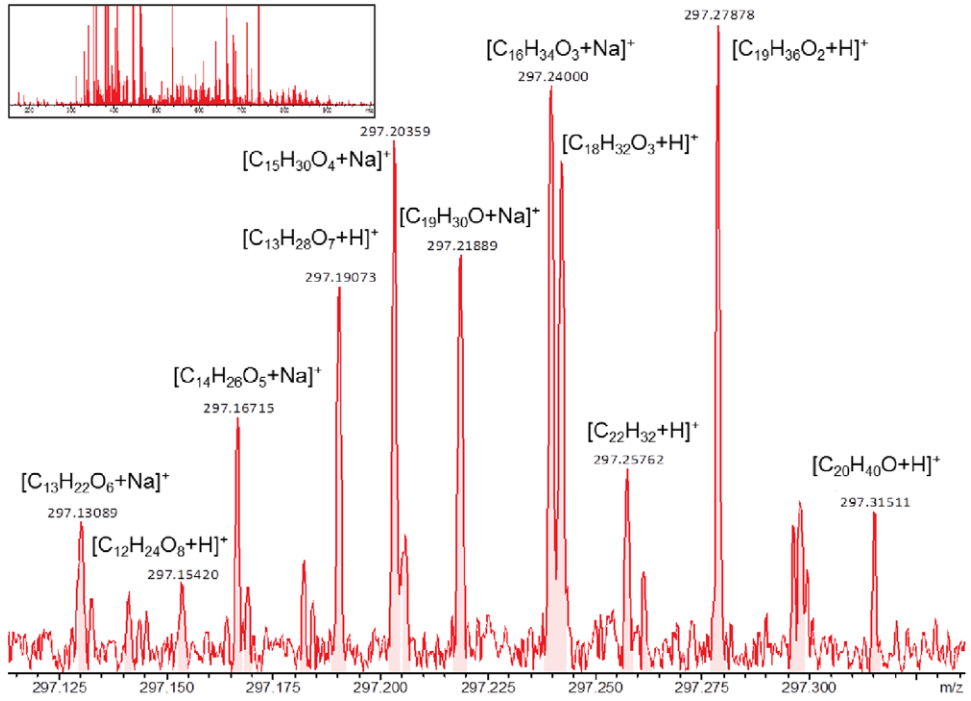

Figure 2. EBC positive mode mass spectrum with > 10 000 *m/z* peaks in top left. Zoom into nominal mass 297 with corresponding ion formula assignments of *m/z* peaks.

For MDEA we used the functions *hygestat* and *hygecdf* that are implemented in Matlab R 2011. Both functions assume a hypergeometric, discontinuous distribution and create different statistics after definition of the following entries: population (here, the entire amount of edges), number of elements of interest in the population (here, the overall frequency of a specific reaction), size of a sample population (here, amount of reactions participating at the interface between VOCs and EBC compositions) as well as 'number of draws' (here, the observed frequency of a single reaction within the sample population).

The significance of enrichment is reported in *p*-values and *z*-scores. *Z*-scores calculate as:

$$z = \frac{x - \mu}{\sigma}, \qquad (1)$$

with x being the observed frequency of a reaction in the sample population, μ being the expected frequency of reactions in the sample population and σ being the standard deviation around μ. *Z*-scores of $z \approx 2$ and $z \approx 2.5$ are proportional to *p*-values of $p \approx 0.05$ and $p \approx 0.01$, respectively.

3. Results

3.1. Annotation of *m/z* features

The EBC-compositional space was calculated on the basis of 23 EBC samples and 4 methanol–water blanks. *m/z* peak counts for EBC spectra ranged between 3136 and 11 045 while blanks were populated with 1075–1429 peaks. All calibrated spectra were aligned into one *m/z*-sample matrix containing 37 996 unique *m/z* signal peaks. Sporadically occurring peaks with an abundance of less than three peaks throughout all samples were removed, which reduced the *m/z* peak count to 10 154. Subsequently, isotopic peaks were identified by screening the *m/z* matrix for typical isotopic mass differences within an error range of 2 ppm. Isotopic peak candidates were removed if their Pearson correlation coefficient with their mono-isotopic peaks was > 0.9. Furthermore, peaks with abnormal mass defects were excluded: the decimal places of compound masses (*m/z* values with $z = 1$) composed of CHNOPS are bounded by the decimal places of each atom of each element it contains. Therefore, no chemical compound of e.g. *m/z* = 100.5 can exist if it is composed of CHNOPS. Such a compound would either represent an ion of *m/z* = 201/2 or it cannot be composed of CHNOPS exclusively. In-house lists of theoretical compounds were used to produce a mass defect over *m/z* map, which indicates allowed ranges. Features that were not found within these bounds were excluded.

Both filters reduced the amount of CHONSP peaks with $z = 1$ down to 8200 peaks. The ionic molecular formulas of those data were annotated using a MDiN approach as published in [37]. The approach annotated 4018 *m/z* peaks, 297 of which were removed because they deviated strongly from the joint error over *m/z* distribution (figure 2).

The remaining 3721 ion annotations contained multiple features of identical composition but different ion type (e.g. $[M + H]^+$ or $[M + Na]^+$). The intensities of those features were summed up, which resulted in an overall amount of 2762 compositions. For further analysis we considered all 2691 compositions (see the

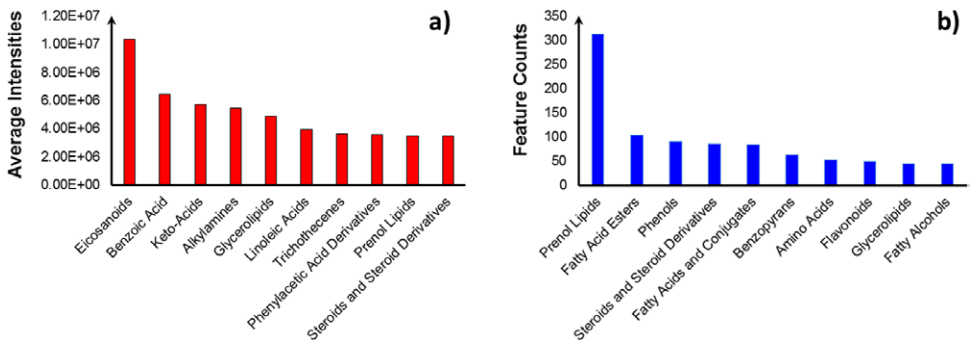

Figure 3. Most abundant calculated compositions annotated to functional compound classes. (*a*) Top 10 compound classes sorted for average ion abundance. (*b*) Top 10 compound classes sorted for feature count.

supplementary data at stacks.iop.org/JBR/9/027105/mmedia) whose average intensity was at least twice as large as the corresponding average blank intensity. 2361 compositions were found in EBC exclusively. Other compositions, such as palmitic acid ions are commonly observed trace contaminations, however, if they are enriched in a sample, they have to be included into the list of compounds that characterize the sample.

3.2. Compositional space

It is now possible to describe the compositional space of EBC. Traditional approaches encompass van Krevelen analysis [38] or Kendrick mass defect analysis [39]. A coarse characterization of the EBC compositional space is given by counting the abundance of different elemental combinations. The given EBC data contained 1451 CHO compositions, 985 CHON compositions, 94 CHN compositions; only 83 and 39 compositions contained sulfur and phosphorus, respectively. Furthermore, EBC can be characterized by exact mass matching to metabolic databases. We compared the given EBC compositions to 8585 compositions derived from the Human Metabolome Database (HMDB), version 3.5 [40]. The database contained 52 140 entries, which pertained to 8585 CHNOPS sum formulas. Each sum formula was attributed with the compound classes of all its isomers and the two most abundant compound classes were used for the generation of figure 3.

With 1095 matching compositions (41%), the proportion of EBC formulas matching to HMDB was exceptionally high (normal frequencies of database annotation are between 5 and 20% using our platform).

The most prominent compound classes according to average compound class ion abundance and compound class feature count are presented in figure 3.

The majority of the 235 VOC compositions were represented by the following compound classes (and abundancies): N-containing compounds (72), S-containing compounds (26), furans (23), alkoxy-ethers (17), aldehydes (16), alkanes (16), cyclic ethers (15), unsaturated aldehydes (13), branched chain alkanes (12), branched dienes (12), alcohols (12), acetic acid esters (12), aliphatic aldehydes (12).

An excellent discussion about these compounds is to be found in [2].

3.3. Mass difference networking

The theoretical neutral masses both of EBC compositions and VOC compositions were translated into an MDiN. Two compositions (nodes) were connected by an edge if they were found to exhibit a mass difference that mimics a biochemical reaction (e.g. acetylation: product − substrate = C_2H_2O = 42.010565 Da; an alternative example is shown in figure 1). All 2691 EBC compositions and 235 VOC compositions were networked together and resulted in an MDiN of 2926 nodes and 115 123 edges. 12 921 edges were found between VOCs and EBC. In general edges are directed from the node with smaller mass to the node with larger mass. We found that the vast majority of VOC compounds was the node of smaller mass if they were connected to an EBC compound—as expected; since VOCs are products of catabolism. One exception is replacement of an amino functionality with a hydroxylic group; a variation of transamination. It was generally the OH containing compound that was found in the VOC, while the N-containing compound of smaller mass was found in EBC. As published in Zhang *et al* [34], we performed MDEA in order to gain information upon the significance of reaction for the interface between EBC metabolome and breath volatilome.

We found that 76 out of 189 reactions was pronounced at the VOC–EBC interface with p-values that withstood Benjamini–Hochberg correction at $\alpha = 0.01$. The reaction classes that are most pronounced in the VOC–EBC interface are condensation and dehydrogenation followed by condensation of carbonyls and primary amines.

In contrast, 39 reactions were strongly under-represented. Surprisingly, methylation, hydrogenation, C2-unit elongation or hydroxylation (p-values: 2.9×10^{-60}, 4.4×10^{-53}, 3.2×10^{-49} and 9.8×10^{-44}, respectively) were found among the most under-represented reactions. While those REMDs occurred between 1700 to 2100 node-pairs throughout the entire MDiN, not

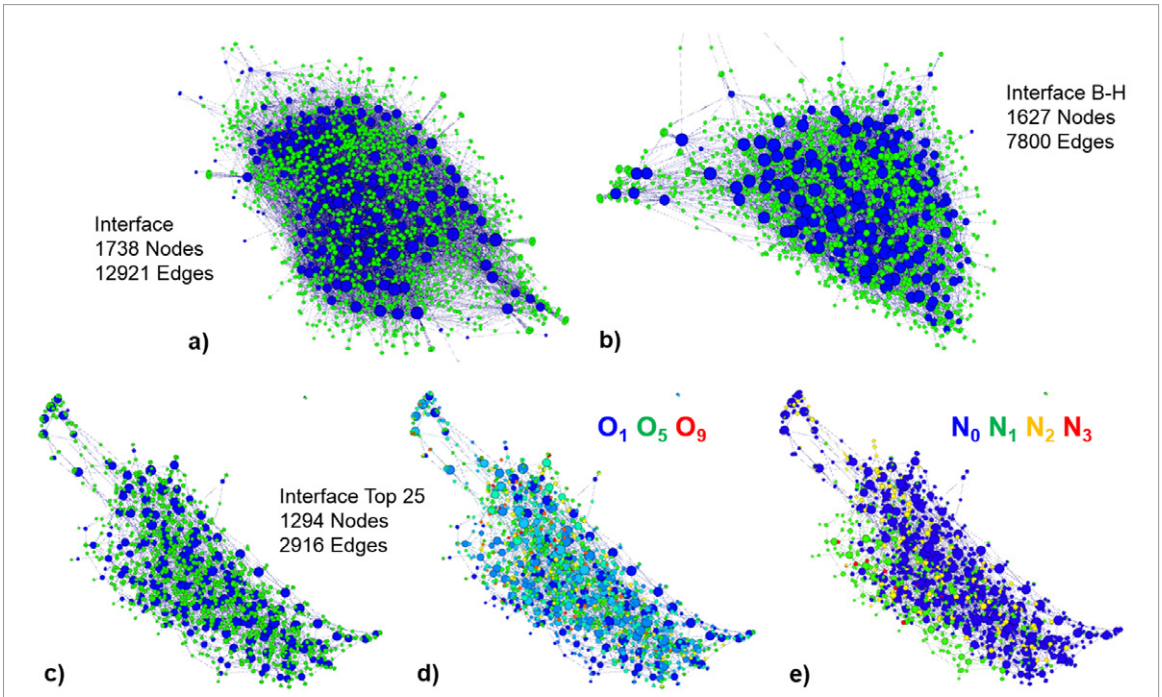

Figure 4. MDiN of the EBC–VOC interface. Nodes are compositions, edges are reaction mass differences. Node sizes are scaled for connectivity (degree). (*a*) full EBC–VOC interface. EBC nodes are green; VOC nodes are blue. (*b*) EBC–VOC interface laid out with reaction mass differences that were significantly enriched after Benjamini–Hochberg correction. VOC nodes are blue, EBC nodes are green. (*c*)–(*e*) EBC–VOC interface laid out using the 25 top-scoring reaction mass differences. Coloring in (*c*) is like coloring in (*a*) and (*b*). Coloring in (*d*) progresses from blue over green to red with increasing oxygen content. Coloring in (*e*) progresses from blue over green to red with increasing nitrogen content.

more than 30–50 pairs per REMD were found at the interface.

Figure 4 demonstrates how the degree of organization of the EBC–VOC interface improves as it is laid out under gradual exclusion of non-enriched reaction mass differences.

Figure 4(*a*) shows the entire detected EBC–VOC interface given all 189 used reactions. Notably, no VOC node is connected to another VOC nodes and the same is true for EBC nodes. The interface network is surprisingly connected.

Furthermore, MDEA was performed for the interactions of VOCs with the compound classes shown in figure 3(*a*). The five most significantly over-represented reaction mass differences for each compound class are summarized in table 1.

The 10 most abundant compound classes found in EBC used an average of 17 significantly enriched (z-scores with $z > 2$ with $p < 0.05$ after Benjamini–Hochberg correction at $\alpha = 0.05$) REMDs to connect to the volatilome. Table 1 contains the top five REMDs for each compound class. Similar compound classes were unified into one column. Z-scores of REMDs that were found to be shared by the compound classes in one column are separated by a vertical line.

4. Discussion

4.1. EBC compound classes

Querying GoPubmed.org for 'exhaled breath condensate' yields a list of more than 1000 publications about the analysis of EBC. Screening the abstracts of all publications, we found that the majority of papers on EBC deal with pH or with inorganic compounds such as H_2O_2, nitric oxide (more than 350 instances; e.g. [41]). These compounds are followed by isoprostanes, prostanoids, leukotrienes and eicosanoids in general (>100 instances; e.g. [42]). Interleukines and other chemokines were found in more than 70 and more than 50 instances, respectively. 19 references investigate 3-nitrotyrosine (e.g. [43]). Some amino acids and many short chain acids were analyzed, but referencing them all would be out of scope at this point, as we screened the abstracts of more than 1000 articles. The overall number of organic target compounds measured is slightly above 100. The focus on these compounds reflects that the major scientific background of EBC investigations lies in the fields of pneumology and immunology. A systemic origin of EBC constituents is not well described in literature. Most metabolomics experiments on EBC were performed using NMR, which should clearly be advantageous in regard to differential dilution of EBC components. The amount of metabolites detected by means of NMR does commonly not exceed 40 or 50 compounds. The vastness of the chemical space in EBC has not been described yet. It is, however, important to know which compound classes are likely to be found in EBC, as this knowledge is fundamental for the formulation of biochemical/medical hypotheses, for the estimation of pathways and sites of breath VOC biosynthesis and for developing ideas concerning which different pathways

Table 1. *Z*-scores for enrichment of reaction-mass differences at the interface between EBC compound classes and VOCs. Reaction classes are: (A) condensation/hydrolysis, (B) dehydrogenation followed by condensation of carbonyls and primary amines, (C) others, (D) decarboxylative condensation.

Reaction partner	Reaction class	Formula difference	Mass difference	Eicosanoids|linoleic acids	Benzoic acids|phenylacetic acids	Prenol lipids|steroids	Keto acids	Alkylamines	Glycerolipids
Sebacic acid	A	$C_{10}H_{16}O_3$	184.109945	11.3|9.29		8.92			
Hexadecanoic acid	A	$C_{16}H_{30}O$	238.229665	8.02					
Quinate	A	$C_7H_{10}O_5$	174.052825	6.37					8.33
4-Hydroxyphenylpyruvic acid	A	$C_9H_6O_3$	162.031695	5.5	6.41	9.81			
Ferulic acid	A	$C_{10}H_8O_3$	176.047345	5.5	3.86|6.03	8.78|13.41			
Azelaic acid	A	$C_9H_{14}O_3$	170.094295	7.46					6.41
Tetradecanoic acid	A	$C_{14}H_{26}O$	210.198365	6.49					
Farnesylation	C	$C_{15}H_{24}$	204.1878	5.54		11.36			
Decanoic acid	A	$C_{10}H_{18}O$	154.135765	5.02		9.16			
Phenylpyruvic acid	A	$C_9H_6O_2$	146.03678		4.98	11.97			
Salicylic	A	$C_7H_4O_2$	120.02113		4.39|3.07				
Indole, pyruvate, ammonia	C	$C_2H_3NO_2$	73.016379		3.86				
2-Ketosuccinate	D	$C_4H_2O_4$	113.99531		3.36				
4-Hydroxyphenylpyruvic acid	A	$C_8H_8O_2$	136.05243		3.84				
Adipate	A	$C_6H_8O_3$	128.047345		2.84				
Dodecanoic acid	C	$C_{12}H_{22}O$	182.167065			8.48			
Geranylation	C	$C_{10}H_{16}$	136.1252			8.37			
Geranylgeranylation	A	$C_{20}H_{32}$	272.2504			9.4			
Hydroxypyruvic acid	A	$C_3H_2O_3$	86.000395				3.31		
3-Hhydroxy-2-oxobutanoic acid			100.016045				2.95		
Glucose	A	$C_6H_{10}O_5$	162.052825				2.69		9.38
Rhamnose	A	$C_6H_{10}O_4$	146.05791				2.61		9.31
Shikimic acid	A	$C_7H_8O_4$	156.04226				2.61		
Serine	A	$C_3H_7NO_2$	89.047679					3.31	
Threonine	A	$C_4H_7NO_2$	101.047679					2.77	
Threonine	B	$C_4H_9NO_2$	103.063329					4.62	
Leucine/isoleucine	B	$C_6H_{13}NO$	115.099714					3.49	
Glutamic acid	B	$C_5H_9NO_3$	131.058244					3.31	
Glucuronidation	A	$C_6H_8O_6$	176.03209						8.91

or actions of different tissues are actually multiplexed into bio-fluidic systems [35, 44].

We propose that a first link between precursor containing bio-fluids and the volatilome could be formed by comparing annotated UHR-MS data to the published volatilome. The compendium published by de Lacy Costello *et al* [2] is therefore an invaluable resource for the broad band comparison of all relevant sample types and 'omes'. While UHR-MS cannot immediately identify and quantify chemical compounds, it provides a wealth of exact compositional information that can readily be used to discern major biochemical trends.

As depicted in figure 3, 1095 of 2691 compositions could be successfully mapped against the HMDB database and the majority of these database annotations (>90%) is covered by 10 compound classes. Noteworthy, only three compound classes—prenol lipids, steroids and glycerolipids—are both among the top 10 most abundant compositions and among the top 10 compound classes in terms of average ion abundance (i.e. that are either very abundant or very sensitive to electrospray ionization). In particular, alkylamines, which include carnitines, are found six times while they are on the fourth rank in terms of ion abundance. Similarly, eicosanoic acids and linoleic acids, which are among the most prominent targets in EBC research, are detected very sensitively while they do not present the major proportion of EBC compositions. Various fatty acid derivatives and amino acid derivatives are found at broad complexity; however, even though they might carry important information on metabolism, they are unlikely to be identified in targeted analyses since their individual ion abundances are low. Low ion abundances do not necessarily imply low concentrations, but such compounds are commonly not suitable for structure elucidation via MS/MS experiments. As compared to our general experience, the unperturbed EBC contains remarkably few compositions that contain sulfur or phosphorus.

4.2. Mass difference enrichment analysis

A further benefit of UHR-MS is that exact mass differences between the plethora of found elemental compositions can be used as proxies for enzymatic reactions. This is because any substrate-product pair of an enzymatic reaction expresses a distinct mass difference, which can be used for data analysis. Similar to gene set enrichment analysis (GSEA) [45], where 'concerted' behavior of genes is analyzed even if individual genes express weak statistics, we can analyze the concerted nature of mass differences even if singular m/z peaks are of low ion abundance. A common topic in breath analysis is oxidative stress, which—if it was the major influence upon the metabolome—would hypothetically result in increased oxygenation ($\Delta m/z = 15.994915$) and de/hydrogenation ($\Delta m/z = 2.01565$). Methylation would result in increased observations of $\Delta m/z = 14.01565$.

We commonly use a set of amino acids, keto acids, fatty acids, dicarboxylic acids as well as conjugation with saccharides and glutathione and various smaller mass differences (overall 189) in order to investigate the relationships between sets of compositions that have been shown to be important for some statistic. Here we used these mass differences to investigate, whether their occurrence is particularly pronounced at the EBC–VOC interface as compared to their overall abundance over all compositions. We commonly express their overrepresentation in z-scores, where $z \approx 2$ indicates p-values of $p \approx 0.05$ and $z \approx 2.5$ indicates p-values of $p \approx 0.01$.

We found extreme p-values both for over-representation and for under-representation of individual mass differences at the EBC–VOC interface.

In general, VOCs were found to have the smaller mass in VOC–EBC pairs. VOCs are often stable products of catabolism [46]. Mass differences that were absolutely under-represented are any mass differences that were related to the reaction types 'decarboxylative condensation' and 'decarboxylative addition'. Given a mass gain (here VOC → EBC), those reactions form C–C bonds under loss of H_2O and CO_2. The under-representation of these mass differences is entirely consistent with general chemistry, as finding them in the EBC → VOC interface would imply 'breaking C-C bonds' while inserting H_2O and CO_2, which is not thermodynamically favored. Furthermore, most mass differences related to C_1-metabolism and reactions with inorganic moieties were under-represented. As the given EBC set was sampled from healthy individuals, it stands to be hypothesized whether (a) oxidative stress indicators were under-represented for this reason, whether (b) oxidative stress in not the major effect that drives EBC composition, or whether (c) oxygenation and hydrogenation are not proper oxidative stress markers.

The four most over-represented mass differences were pertained to glutamine and asparagine, which implies their systemic importance in the lung [47]. Glutamic acid and to a lesser extent aspartic acid were observed to be characteristic for potential VOC precursors as well. Glutamine and glutamic acid are vital for lung function and energy metabolism. Also their concentrations in blood are commonly high and they are often found to be conjugated to other molecules. Furthermore, there are specific glutamine–glutamate transporters that mediate the transition from blood to lung [48]. Deconjugation of lysine, ornithine, threonine and other amino acids was pronounced. Another class of compounds whose cleavage was pronounced at the EBC → VOC interface were terpenoid units, hexoses, hexadecanoic and tetradecanoic acid as well as dicarboxylic acids and—surprisingly—ferulic acid. Glucuronidation and glycosylation are known phase II metabolic events [49]. As the enriched mass differences can largely be interpreted to be phase II reactions, the lung attains more and more the nature of an excretion organ. A loss of terpenoid functionalities can often be

associated with exogenous sources of the related precursors [50]; however, as prenol lipids were found to be the most abundant compound class in EBC, much about their nature and role is yet to be uncovered. Hexadecanoic and tetradecanoic acid cleavage may largely be explainable by their cleavage from (phospho-)glycero-lipids. Finally, dicarboxylic acids are mostly considered to be plant metabolites, however in plants they are products of the oxilipin pathway, which produces carbonyls and dicarboxylic acids from linoleic acid and other poly-unsaturated acids [51]. As eicosanoic acids (and arachidonic acids) [52] and linoleic acids [53] were found to be of high individual abundance in EBC, dicarboxylic acid related mass differences may be alternative indicators of oxidative stress. There is a strong tenor on the decomposition of typical phase II products for which reason airway lining fluid may be a significant site of VOC production.

Figure 4 shows different lay-outs of the EBC–VOC interface. Laying out the full EBC–VOC interface with 1738 nodes (235 VOC; 1503 EBC) and 12 921 edges, we expected the occurrence of many small and disconnected networks. However, we found tightly interwoven compositional relationships between VOCs and EBC. It is to be noted, that only 18 VOC compositions were found in EBC themselves. Those 18 VOC compositions majorly encompassed 39 single VOCs; among them 10 allyl/alkenyl ethers, six alcohols, four acetic acid esters and three aliphatic primary acids. Fourteen VOC compositions contained enough oxygen to formulate carboxylic functionalities. The found compositions that are shared between VOCs and EBC are highlighted in supplementary table S1 (stacks.iop.org/JBR/9/027105/mmedia). 56% of all EBC compositions were found at the interface. This tightly interwoven network structure underlines the fundamental centrality of VOC compositions in the compositional space. It is to be noted, that VOC compositions generally exhibited a median degree (number of connections) of 52 with EBC, while EBC nodes exhibited a median degree of 6 with VOCs. This result implies a strong participation or at least a strong compositional relationship between both compositional spaces. This is only natural, as VOCs can likewise be addressed as building blocks of precursors. After Benjamini–Hochberg correction at $\alpha = 0.01$ we decreased the amount of reactions for the lay-out down to 73. The laid out network was still highly connected with 7800 edges and with 52% of all EBC compositions participating in the interface (figure 4(*b*)). We further reduced the amount of used reactions down to 25, which resulted in a network of 2916 edges and 1294 node (39% of all EBC compositions). Visual network inspection of figure 4(*c*) reveals a more ordered network pattern: alternating line-ups of VOCs and EBCs. These stripe-like structures are homologous series given the used reactions. Figure 4(*d*) is colored for oxygen content. It shows a clear separation of degrees of oxygenation with EBC compounds being more oxygen rich than VOC compounds. Inspection of figure 4(*e*), which is colored for nitrogen content, shows a more clear segmentation. This strong separation of nitrogen levels, as compared to the separation of oxygen levels, pronounces the potential of phase II reaction products as precursors. The structure of the network is built so as to only connect nodes that are not from the same compositional space (VOC and EBC). This adds significance to the clear spatial separation of nitrogen compounds in figure 4(*e*), as this means that only compounds which are comparably poor in nitrogen are precursors to breath VOCs. Condensed N-aromatics as they occur in nucleotides, as well as peptides or compounds containing guanidine groups seem not to participate in breath VOC generation.

In order to gain more information about the nature of over-represented mass differences, we performed enrichment analyses on the interfaces between the EBC compound classes of highest ion abundance and VOC

4.3. Compound class specific MDEAs

4.3.1. Eicosanoic acids and linoleic acids

Table 1 shows particularly high *z*-scores for the cleavage of sebacic acid, hexadecanoic acid, quinate, hydroxyphenylpyruvic acid and ferulic acid for eicosanoic acids. We explained the relationship between dicarboxylic acids and oxilipin metabolism in the last section, for which reason it makes sense to find sebacic acid at the top of the list of enriched reactions for eicosanoic acids and linoleic acids. Cleavage of hexadecanoic acid may either merely imply the compositional relationship between eicosanoic acids and C_4-VOCs, or it may indeed imply the systematic loss of $C_{16}H_{30}O$-moieties. Quinate, a sugar acid commonly found in plants [54] may pronounce the relationship between oxygenized eicosanoids and e.g. C_{14}-alkanes, alkenes or alcohols and ethers. Both, hydroxyphenylpyruvic acid and ferulic acid mass differences (a plant metabolite) do most likely not express the loss of actually these moieties, but they may indicate a systematic compositional relationship of oxygenated arachidonic acids and C_{10} to C_{11} alkanes, or aliphatic alcohols. Reactions that further characterized linoleic acids were found to be azelaic acid (again a dicarboxylic acid), tetradecanoic acid, decanoic acid and cleavage of a farnesyl moiety ($C_{15}H_{24}$). Removing $C_{15}H_{24}$ from linoleic acid gives $C_3H_8O_2$, a propanediol or ether moiety. In this case, farnesyl cleavage does likely not address the actual cleavage of $C_{15}H_{24}$.

4.3.2. Benzoic acids and phenylacetic acids

Benzoic acids [55] and phenylacetic acids [56] are both secondary plant metabolites. There occurrence and type of conjugation in EBC might allow insight in a person's nutritional and intestinal status [57].

Except for cleavage of adipate, all reactions pertaining to both EBC compound classes indicate a loss of some aromatic moiety. It is very well plausible that the cleavage of relevant esters leads to the production of VOCs. Adipate, a C_6-dicarboxylic acid ($C_6H_8O_4$) has a

degree of hydrogen deficiency similar to that of benzoic acids and phenylacetic acids. This reaction candidate may therefore indicate compositional relationships between poly-substituted aromatics (e.g. with O–Me or O–Et) and aliphatics among the VOCs.

4.3.3. *Prenol lipids and steroids*

Prenols and steroids in breath and EBC are poorly described. However, given their abundance in EBC, clarification about their origins is necessary.

Here, prenol lipids are characterized by de-geranylation while Steroids are characterized by de-farnesylation and de-geranylgeranylation. In both cases, those findings are consistent with the terpenoid structure of individual prenols or steroids. The fact that only one terpenoid backbone related reaction is found for prenol lipids, derives from the fact that prenols are the most common compound class in EBC, which makes those reactions a common event. As MDEA indicates over-representations of reactions/mass differences, only geranyl-units seem to be abnormally enriched. It is likewise the most probably unit, as most VOCs do not contain more than 10 carbon atoms and adding $C_{10}H_{16}$ localizes potential precursors at ≈300 Da, which is the most populated mass region in EBC. The remaining mass differences that are over-represented in table 1, again dicarboxylic acids and aromatic acids connect EBC–prenols and EBC–steroids to rather aliphatic VOCs that are poor in oxygen content. Decanoic and dodecanoic acid connect prenols to rather unsaturated VOCs.

4.3.4. *Keto-acids*

Keto-acids are important substrates and products of fatty acid metabolism, and play important roles in central carbon and nitrogen metabolism [58]. They are furthermore produced via transamination from amino acids in the liver. We often observe these compositions in EBC but there is poor knowledge especially about α-keto acids. In the HuMet study, we observed keto-acids to stand in a relation to blood carnitine and acetone levels [13]. In depth investigations pertaining to the role of α-keto acids in blood would be very interesting and might give insights upon a patient's liver-status.

Keto-acids themselves have a rather high degree of unsaturation as well as high oxygen content, for which reason their most pronounced reactions pertain to hydroxypyruvic acid, 2-hydroxy-2-oxobutanoic acid (the keto-acid of serine), glucose, rhamnose and shikimic acid. The first two mass differences can be potential leaving groups that actively produce VOCs, while the latter three mass differences are likely to be compositional analogies.

4.3.5. *Alkylamines*

Alkylamines detected in EBC cover carnitines, ethanolamines and panthothenols. Carnitines are necessary to transport fatty acids through membranes. We have found more carnitine candidates in not yet published studies; they potentially enable surveillance of the nutritional state (e.g. fastening versus oral glucose tolerance tests) [13]. The ethanolamines found here (diethanolamine and triethanolamine) indicate contamination from cosmetics or cleaning agents. Panthenols are support the function of mucosa and skin function, for which reason they are often added to cosmetics [59]. It might be possible that these compounds are endogenous, which would be of special interest for the diagnostics of pulmonary function.

Here, we found that alkylamines are connected to VOC via cleavage of serine, threonine, leucine/isoleucine and glutamic acid. Hydrolyses of these compositions from the given alkylamines might indeed release alkanes, alkenes and hydroxy-alkanes. Lysine is one of the precursors for carnitines and all compositions may address different concerted losses of functionalities that end up in VOCs.

4.3.6. *Glycerolipids*

Glycerolipids are constituents of a multitude of membranes and depending on their degree of acylation, they may act as mediators for metabolic processes or may just support mechanical lung function. Most mass differences that were over-represented for these compounds pertain to the homology between acyclic alkanes, alkenes, carbonyls or acids from VOC and the acyl-residues of glycerolipids. For this reason, the most observed reactions are quinate, azelaic acid, glucose, rhamnose and glucuronidation. They all pertain to the glycerol-acyl-ester moieties contained in glycerolipids. The removal of these compositions may realistically produce VOCs.

The mass differences which were enriched in the 'global' MDEA–amino acid based reactions are barely found among the analyses of the special cases. Even though the HMDB-annotation covered 41% of all annotated features, there are 59% which are not contained by this database. It is to be noted that exact mass differences are not guaranteed to reflect the exact reactions which they were meant to mimic. While there are a multitude of composition pairs which are not likely to display a real reaction, real metabolic pathways are guaranteed to be a subset of the MDiN as long as the respective metabolites participating in them were detected and as long as the relevant mass differences were addressed. We analyzed the known metabolic KEGG maps and found that the entire primary metabolism can be mimicked by fewer than 15 reactions. The comparative analysis of EBC compositional space and the human volatilome as proposed here is one of many possibilities to mine information about chemical relationships at system borders. Simple graph theoretical path search algorithms such as Dijkstra's algorithm can be used to mine pathway candidates between a targeted VOC and a source–EBC composition. The power of these approaches increases with experimental context. Here, our scope was to present the general 'unperturbed' compositional space of EBC. However,

experiments that are properly designed to investigate, e.g., diseases lead to can lead to MDEA results which pronounce relevant sections of the given compositional space. Such a dynamic scenario increases the power of this approach as it allows us to track pathways as they change in response to the experiment. The optimum scenario would be a sample set (and finally a data set) where EBC and VOCs were measured in parallel. At the moment no such apparatus exists, but it is worth investigating and developing in this direction. A method, which might bridge or combine both investigated compositional spaces is secondary electrospray ionization (SESI) mass spectrometry [60], as this technique might potentially detect both VOCs and non-VOCs. Another possibility would be the analysis of EBC headspace (after it EBC has been sampled) and subsequent UHR-MS analysis of EBC itself. It is even imaginable that such a strategy would provide valuable information for adjusting EBC dilution.

In order to effectively tackle the task of VOC-generating pathway identification, it is important for both scientific disciplines, the highly quantitative analysis of breath VOCs and the semi-quantitative UHR-MS analysis of bio-fluids (here EBC), to integrate their efforts. Both worlds are not easily unified, as conventional quality measures of instrumental analysis, such as limit of quantitation and coefficient of variation, are not adequate indicators of quality in non-targeted broad scan measurements. Rather than univariate 'standards', multivariate behaviors, concerted action of groups of analytes, supported by broad band mass accuracy are measures of interest in explorative UHR-MS. Among many other sample matrices, our laboratory handles data of blood plasma, feces, urine and saliva and we will gladly take up the path opened by de Lacy Costello *et al* to investigate precursor candidates for these bodily fluids.

5. Conclusion

We have shown that EBC harbors thousands of compositions (and even more individual chemical compounds), and that the proportion of compound classes does not reflect the frequency by which individual compounds have been targeted so far. Indeed, we have found that eicosanoids and linoleic acids are among the groups with highest ion abundance, however, excluding prenol lipids (which were not investigated in EBC yet) the top nine most intense compound classes cover but 15% of all possible annotations. With MDEA we have proposed a way to track VOC-precursor candidates in EBC and have shown that phase II reaction related mass differences are very much more specific for the EBC–VOC interface than oxidative stress indicators. In conclusion there is a potential for the airway lining fluid to be a significant site of VOC production. This assumption is supported by the richness of an MDiN that was reconstructed on the basis of VOCs. We have outlined that coordinated VOC–EBC sampling from phenotypes of interest may be the way to go in order to achieve an effective exploration of VOC-generating pathways.

Acknowledgments

This work was supported and funded by the German Center for Diabetes Research and by the Polish Committee for Scientific Research (research project #N N404 201639).

References

[1] Amann A, Costello Bde L, Miekisch W, Schubert J, Buszewski B, Pleil J, Ratcliffe N and Risby T 2014 The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva *J. Breath Res.* **8** 034001

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

[3] Walker A *et al* 2014 Distinct signatures of host-microbial meta-metabolome and gut microbiome in two C57BL/6 strains under high-fat diet *ISME J.* **8** 2380–96

[4] Stobbe M D, Houten S M, Jansen G A, van Kampen A H and Moerland P D 2011 Critical assessment of human metabolic pathway databases: a stepping stone for future integration *BMC Syst. Biol.* **5** 165

[5] Erhart S *et al* 2009 3-Heptanone as a potential new marker for valproic acid therapy *J. Breath Res.* **3** 016004

[6] Effros R M 2010 Exhaled breath condensate: delusion or dilution? *Chest* **138** 471–2

[7] Janicka M, Kot-Wasik A, Kot J and Namiesnik J 2010 Isoprostanes-biomarkers of lipid peroxidation: their utility in evaluating oxidative stress and analysis *Int. J. Mol. Sci.* **11** 4631–59

[8] Yuksel H, Turkeli A, Taneli F, Horasan G D, Kanik E T, Kizilkaya M, Gozukara C and Yilmaz O 2014 E-cadherin as an epithelial barrier protein in exhaled breath condensate *J. Breath Res.* **8** 046006

[9] Xiao P, Chen J R, Zhou F, Lu C X, Yang Q, Tao G H, Tao Y J and Chen J L 2014 Methylation of P16 in exhaled breath condensate for diagnosis of non-small cell lung cancer *Lung Cancer* **83** 56–60

[10] Janicka M, Kubica P, Kot-Wasik A, Kot J and Namiesnik J 2012 Sensitive determination of isoprostanes in exhaled breath condensate samples with use of liquid chromatography-tandem mass spectrometry *J. Chromatogr.* B **893–894** 144–9

[11] Nicholson J K 2006 Global systems biology, personalized medicine and molecular epidemiology *Mol. Syst. Biol.* **2** 52

[12] Carraro S, Giordano G, Reniero F, Carpi D, Stocchero M, Sterk P J and Baraldi E 2013 Asthma severity in childhood and metabolomic profiling of breath condensate *Allergy* **68** 110–7

[13] Krug S *et al* 2012 The dynamic range of the human metabolome revealed by challenges *FASEB J.* **26** 2607–19

[14] Romano A, Aiani M R, Brambilla M, Brambilla P, Carcano M, Galbiati E, Porro S and Mascagni P 2012 Metabolomic profiles of exhaled breath condensate of 39 nickel exposed workers *G. Ital. Med. Lav. Ergon.* **34** 682–6

[15] Li C *et al* 2014 Estradiol and mTORC2 cooperate to enhance prostaglandin biosynthesis and tumorigenesis in TSC2-deficient LAM cells *J. Exp. Med.* **211** 15–28

[16] Baraldi E, Carraro S, Giordano G, Reniero F, Perilongo G and Zacchello F 2009 Metabolomics: moving towards personalized medicine *Ital. J. Pediatr.* **35** 30

[17] Carraro S, Rezzi S, Reniero F, Heberger K, Giordano G, Zanconato S, Guillou C and Baraldi E 2007 Metabolomics applied to exhaled breath condensate in childhood asthma *Am. J. Respir. Crit. Care Med.* **175** 986–90

[18] de Laurentiis G, Paris D, Melck D, Maniscalco M, Marsico S, Corso G, Motta A and Sofia M 2008 Metabonomic analysis

of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy *Eur. Respir. J.* **32** 1175–83
[19] de Laurentiis G, Paris D, Melck D, Montuschi P, Maniscalco M, Bianco A, Sofia M and Motta A 2013 Separating smoking-related diseases using NMR-based metabolomics of exhaled breath condensate *J. Proteome Res.* **12** 1502–11
[20] Ibrahim B, Marsden P, Smith J A, Custovic A, Nilsson M and Fowler S J 2013 Breath metabolomic profiling by nuclear magnetic resonance spectroscopy in asthma *Allergy* **68** 1050–6
[21] Izquierdo-Garcia J L, Peces-Barba G, Heili S, Diaz R, Want E and Ruiz-Cabello J 2011 Is NMR-based metabolomic analysis of exhaled breath condensate accurate? *Eur. Respir. J.* **37** 468–70
[22] Montuschi P, Paris D, Melck D, Lucidi V, Ciabattoni G, Raia V, Calabrese C, Bush A, Barnes P J and Motta A 2012 NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis *Thorax* **67** 222–8
[23] Motta A, Paris D, D'Amato M, Melck D, Calabrese C, Vitale C, Stanziola A A, Corso G, Sofia M and Maniscalco M 2014 NMR metabolomic analysis of exhaled breath condensate of asthmatic patients at two different temperatures *J. Proteome Res.* **13** 6107–20
[24] Motta A, Paris D, Melck D, de Laurentiis G, Maniscalco M, Sofia M and Montuschi P 2012 Nuclear magnetic resonance-based metabolomics of exhaled breath condensate: methodological aspects *Eur. Respir. J.* **39** 498–500
[25] Sinha A, Krishnan V, Sethi T, Roy S, Ghosh B, Lodha R, Kabra S and Agrawal A 2012 Metabolomic signatures in nuclear magnetic resonance spectra of exhaled breath condensate identify asthma *Eur. Respir. J.* **39** 500–2
[26] Sofia M, Maniscalco M, de Laurentiis G, Paris D, Melck D and Motta A 2011 Exploring airway diseases by NMR-based metabonomics: a review of application to exhaled breath condensate *J. Biomed. Biotechnol.* **2011** 403260
[27] Carpagnano G E 2011 Exhaled breath analysis and sleep *J. Clin. Sleep Med.* **7** S34–7
[28] Hassanein M, Callison J C, Callaway-Lane C, Aldrich M C, Grogan E L and Massion P P 2012 The state of molecular biomarkers for the early detection of lung cancer *Cancer Prev. Res. (Phila)* **5** 992–1006
[29] Leung T F, Ko F W and Wong G W 2013 Recent advances in asthma biomarker research *Ther. Adv. Respir. Dis.* **7** 297–308
[30] Maniscalco M, Palladino F, Mormile M and Sofia M 2009 Exhaled nitric oxide and other major exhaled compounds for the diagnosis of metabolic diseases *Expert Opin. Med. Diagn.* **3** 547–56
[31] van Mastrigt E, de Jongste J C and Pijnenburg M W 2014 The analysis of volatile organic compounds in exhaled breath and biomarkers in exhaled breath condensate in children—clinical tools or scientific toys? *Clin. Exp. Allergy* doi: 10.1111/cea.12454
[32] Koch B P, Dittmar T, Witt M and Kattner G 2007 Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter *Anal. Chem.* **79** 1758–63
[33] Schmitt-Kopplin P *et al* 2012 Dissolved organic matter in sea spray: a transfer study from marine surface water to aerosols *Biogeosciences* **9** 1571–82
[34] Zhang F, Harir M, Moritz F, Zhang J, Witting M, Wu Y, Schmitt-Kopplin P, Fekete A, Gaspar A and Hertkorn N 2014 Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry *Water Res.* **57** 280–94
[35] Lucio M *et al* 2010 Insulin sensitivity is reflected by characteristic metabolic fingerprints—a Fourier transform mass spectrometric non-targeted metabolomics approach *PLoS One* **5** e13317
[36] Breitling R, Ritchie S, Goodenowe D, Stewart M L and Barrett M P 2006 Prediction of metabolic networks using Fourier transform mass spectrometry data *Metabolomics* **2** 155–64
[37] Tziotis D, Hertkorn N and Schmitt-Kopplin P 2011 Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity *Eur. J. Mass Spectrom.* **17** 415–21
[38] Hertkorn N, Ruecker C, Meringer M, Gugisch R, Frommberger M, Perdue E M, Witt M and Schmitt-Kopplin P 2007 High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems *Anal. Bioanal. Chem.* **389** 1311–27
[39] Herzsprung P, Hertkorn N, von Tumpling W, Harir M, Friese K and Schmitt-Kopplin P 2014 Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view *Anal. Bioanal. Chem.* **406** 7977–87
[40] Wishart D S *et al* 2007 HMDB: the Human Metabolome Database *Nucl. Acids Res.* **35** D521–6
[41] Taraldsoy T, Bolann B J and Thorsen E 2007 Reduced nitric oxide concentration in exhaled gas after exposure to hyperbaric hyperoxia *Undersea Hyperb. Med.* **34** 321–7
[42] Caballero S, Martorell A, Escribano A and Belda J 2013 Markers of airway inflammation in the exhaled breath condensate of preschool wheezers *J. Investig. Allergol. Clin. Immunol.* **23** 7–13
[43] Lee J S, Shin J H, Hwang J H, Baek J E and Choi B S 2014 Malondialdehyde and 3-nitrotyrosine in exhaled breath condensate in retired elderly coal miners with chronic obstructive pulmonary disease *Saf. Health Work* **5** 91–6
[44] Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, Tysk C and Schmitt-Kopplin P 2009 Metabolomics reveals metabolic biomarkers of Crohn's disease *PLoS One* **4** e6386
[45] Subramanian A *et al* 2005 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles *Proc. Natl Acad. Sci. USA* **102** 15545–50
[46] Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J and Cataneo R N 1999 Variation in volatile organic compounds in the breath of normal humans *J. Chromatogr.* B **729** 75–88
[47] Balzola F A and Boggio-Bertinet D 1996 The metabolic role of glutamine *Minerva Gastroenterol. Dietol.* **42** 17–26
[48] Bode B P 2001 Recent molecular advances in mammalian glutamine transport *J. Nutr.* **131** 2475S–85S
[49] Nakajima M and Yokoi T 2005 Interindividual variability in nicotine metabolism: C-oxidation and glucuronidation *Drug Metab. Pharmacokinet.* **20** 227–35
[50] Braga S F and Almgren M M 2013 Complementary therapies in cystic fibrosis: nutritional supplements and herbal products *J. Pharm. Pract.* **26** 14–7
[51] Matsui K 2006 Green leaf volatiles: hydroperoxide lyase pathway of oxylipin metabolism *Curr. Opin. Plant Biol.* **9** 274–80
[52] Liang Y, Yeligar S M and Brown L A S 2012 Exhaled breath condensate: a promising source for biomarkers of lung disease *Sci. World J.*
[53] Nording M L, Yang J, Hegedus C M, Bhushan A, Kenyon N J, Davis C E and Hammock B D 2010 Endogenous levels of five fatty acid metabolites in exhaled breath condensate to monitor asthma by high-performance liquid chromatography: electrospray tandem mass spectrometry *IEEE Sensors J.* **10** 123–30
[54] Minamikawa T 1976 A comparative study on the metabolism of quinic and shikimic acids in plants *Bot. Mag.* **89** 141–4
[55] Qualley A V, Widhalm J R, Adebesin F, Kish C M and Dudareva N 2012 Completion of the core beta-oxidative pathway of benzoic acid biosynthesis in plants *Proc. Natl Acad. Sci. USA* **109** 16383–8
[56] Wightman F and Lighty D L 1982 Identification of phenylacetic acid as a natural auxin in the shoots of higher plants *Physiol. Plantarum* **55** 17–24

[57] Cueva C, Moreno-Arribas M V, Martin-Alvarez P J, Bills G, Vicente M F, Basilio A, Rivas C L, Requena T, Rodriguez J M and Bartolome B 2010 Antimicrobial activity of phenolic acids against commensal, probiotic and pathogenic bacteria *Res. Microbiol.* **161** 372–82

[58] Doucette C D, Schwab D J, Wingreen N S and Rabinowitz J D 2011 α-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition *Nat. Chem. Biol.* **7** 894–901

[59] Klocker N, Rudolph P and Verse T 2004 Evaluation of protective and therapeutic effects of dexpanthenol on nasal decongestants and preservatives: results of cytotoxic studies *in vitro Am. J. Rhinol.* **18** 315–20

[60] Martinez-Lozano Sinues P, Kohler M and Zenobi R 2013 Human breath analysis may support the existence of individual metabolic phenotypes *PLoS One* **8** e59909