



High-quality identification of volatile organic compounds (VOCs) originating from breath

Wisnave Arulvasan¹ · Hsuan Chou¹ · Julia Greenwood¹ · Madeleine L. Ball¹ · Owen Birch¹ · Simon Coplowe¹ · Patrick Gordon¹ · Andreea Ratiu¹ · Elizabeth Lam¹ · Ace Hatch¹ · Monika Szkatulska¹ · Steven Levett¹ · Ella Mead¹ · Chloe Charlton-Peel¹ · Louise Nicholson-Scott¹ · Shane Swann¹ · Frederik-Jan van Schooten² · Billy Boyle¹ · Max Allsworth¹

Received: 12 June 2024 / Accepted: 10 August 2024
© The Author(s) 2024

Abstract

Introduction Volatile organic compounds (VOCs) can arise from underlying metabolism and are detectable in exhaled breath, therefore offer a promising route to non-invasive diagnostics. Robust, precise, and repeatable breath measurement platforms able to identify VOCs in breath distinguishable from background contaminants are needed for the confident discovery of breath-based biomarkers.

Objectives To build a reliable breath collection and analysis method that can produce a comprehensive list of known VOCs in the breath of a heterogeneous human population.

Methods The analysis cohort consisted of 90 pairs of breath and background samples collected from a heterogeneous population. Owlstone Medical's Breath Biopsy[®] OMNI[®] platform, consisting of sample collection, TD-GC-MS analysis and feature extraction was utilized. VOCs were determined to be "on-breath" if they met at least one of three pre-defined metrics compared to paired background samples. On-breath VOCs were identified via comparison against purified chemical standards, using retention indexing and high-resolution accurate mass spectral matching.

Results 1471 VOCs were present in >80% of samples (breath and background), and 585 were on-breath by at least one metric. Of these, 148 have been identified covering a broad range of chemical classes.

Conclusions A robust breath collection and relative-quantitative analysis method has been developed, producing a list of 148 on-breath VOCs, identified using purified chemical standards in a heterogeneous population. Providing confirmed VOC identities that are genuinely breath-borne will facilitate future biomarker discovery and subsequent biomarker validation in clinical studies. Additionally, this list of VOCs can be used to facilitate cross-study data comparisons for improved standardization.

Keywords Volatile organic compounds (VOCs) · Breath analysis · Metabolome · Volatile metabolites · Non-invasive biomarkers · Breathomics

1 Introduction

Untargeted metabolomics based on mass spectrometry (MS) can provide insight into human health and disease that may not be apparent using nucleic acid or protein-based analytical approaches (Babu & Snyder, 2023). While metabolomic studies primarily focus on aqueous metabolites in samples such as blood, urine, and feces, breath is a rich and diverse matrix containing thousands of different volatile organic compounds (VOCs) (Costello et al., 2014; Haworth et al., 2022). The non-invasive nature of breath sampling makes it

✉ Wisnave Arulvasan
wisnave.arulvasan@owlstone.co.uk

¹ Owlstone Medical Ltd, Cambridge, UK

² Faculty of Health, Medicine and Life Sciences,
Pharmacology and Toxicology, Maastricht University,
Maastricht, Netherlands

particularly attractive for clinical applications, such as early diagnosis and ongoing longitudinal monitoring.

However, the validation of clinically useful breath biomarkers remains limited. This is likely due, at least in part, to the lack of consistent methodologies and quality controls across the breath research literature (Issitt et al., 2022; Jia et al., 2019). To advance the field of breath analysis, there is an urgent need to develop a robust platform that can accurately identify the VOCs considered to be genuinely originating from the breath (which are comprised of endogenous VOCs derived from metabolic processes and exogenous VOCs, such as microbiome or dietary compounds). These breath-borne VOCs need to be distinguished from background VOCs that arise from the sampling equipment and surrounding air that was inhaled immediately before sampling, which are unrelated to underlying physiology. Establishing an accurate and repeatable methodology will expedite the identification and validation of VOC biomarkers of disease in future studies.

There are multiple approaches for collecting and analyzing breath, each with different advantages, limitations, and challenges (Haworth et al., 2022). Untargeted breath biomarker discovery workflows are most common, which often produce data on unknown VOCs or VOCs tentatively identified by comparison with publicly available standard libraries, such as those provided by the National Institute of Standards and Technology (NIST). This increases the risk of misidentifications due to differences in methodology and instrumentation between untargeted datasets and reference libraries, impeding the replication and validation of findings. Accurate identification of the VOCs in a breath sample requires a comparison to purified chemical standards analyzed using the same instrumentation and methods (Fiehn et al., 2007; Sumner et al., 2007). At least one unique chemical standard is required for every VOC to be identified, but as this is costly and time-consuming, many studies forego this critical process.

Another unmet need in the breath field is standardized methods for background correction (Herbig & Beauchamp, 2014). Many reports have noted the significance of background contributions to the VOCs observed in breath samples, which can originate from multiple potential sources such as ambient air or from breath sample collection equipment (Di Gilio et al., 2020; Westphal et al., 2022). Common background correction techniques in breath analysis include the calculation of an alveolar gradient to identify VOCs that are more abundant in breath (Phillips, 1997), or using a lung washout with synthetic air to identify VOCs likely to be contributed by inhaled background (Hewitt et al., 2022; Maurer et al., 2014; Schubert et al., 2005; Spaněl et al., 2013; Westhoff et al., 2022). However, the success of background correction relies heavily on the quality of the background

measurement. One commonly used method for background measurements is to take a sample of ambient air in the same location where breath sampling is being performed, which neglects compounds originating from the sampling equipment. Since VOCs are ubiquitous in the environment and therefore can be introduced through multiple components and points throughout the analytical process (Pham et al., 2023), it is important to ensure comparable collection and handling of breath and background samples, of which there is already a method that has been previously described for the ReCIVA[®] breath Sampler (Di Gilio et al., 2020; Doran et al., 2017).

Other attempts have been made to develop a compendium of breath biomarkers (Drabińska et al., 2021; Kuo et al., 2020) which are useful assemblies and distillations of the important literature in this field. However, they also have the same limitations as the underlying literature. A general lack of standardization in sampling, analysis, and identification of VOCs means that it is difficult to quickly assign confidence to any single observation without reviewing the underlying literature. In this study, we present a novel methodology that combines robust breath and background collection, analytical distinguishing breath VOCs from background contamination, and VOC identification against chemical standards. We demonstrate the capability of this method by presenting a list of high-confidence breath VOCs identified from a heterogeneous human population.

2 Methods

2.1 Study design and subjects

This observational study was approved by the Reading Independent Ethics Committee RIEC: 290620-1, all participants provided written informed consent. Adults (≥ 18 years; Cambridge, UK) recruited all met the inclusion criteria, were free of active respiratory infection symptoms or diagnoses (including COVID-19) and fasted for at least two hours prior to breath sampling. We also decided to include some volunteers with various chronic diseases to account for potential normal variation in the population, and ensure breadth of VOC detection. All subjects were treated as a single cohort in statistical analysis as the study's intention was not to compare differences between disease and control. Breath samples were collected from 99 adult volunteers between January and February 2022. Nine samples were excluded due to saliva contamination (determined by observation of saliva/bubbles within the tube) ($n=8$) or incomplete collection volume ($n=1$). The final analysis consisted of 90 breath samples and 90 paired system backgrounds (Table 1). Of the 90 adult subjects, 24 had some type of

Table 1 Study cohort demographics. A breakdown of the total sample sizes (n) is shown along the top row, split into columns by the overall cohort, and then divided by sex (male and female). The mean age, BMI, and standard deviation (SD) for the cohort are shown in the top two rows. The total n and different percentages (%) of the cohort for smoking status and ethnicity are shown in subsequent rows, with finer breakdowns outlined

Variable	Overall	Female	Male	P-value
Total N	90	48	42	N/A
Age, mean (SD)	44.0 (16.6)	44.1 (16.7)	44.0 (16.6)	0.961
BMI, mean (SD)	26.2 (6.0)	26.6 (7.2)	25.8 (4.2)	0.515
Smoking Status, n (%)				
Current Smoker	5 (5.5%)	2 (4.2%)	3 (7.2%)	0.521
Ex-Smoker	21 (23.3%)	11 (22.9%)	10 (23.8%)	
Never Smoker	64 (71.1%)	35 (72.9%)	29 (69.0%)	
Ethnicity, n (%)				0.900
Caucasian	62 (68.9%)	33 (68.8%)	29 (69.0%)	
South Asian	12 (13.3%)	6 (12.5%)	6 (14.3%)	
East Asian	9 (10.0%)	6 (12.5%)	3 (7.1%)	
Black	5 (5.6%)	2 (4.2%)	3 (7.1%)	
Mixed Race	2 (2.2%)	1 (2.1%)	1 (2.4%)	

chronic disease including type 2 diabetes, high blood pressure, arthritis, and irritable bowel syndrome.

2.2 Breath sampling and analysis

The methodology utilized as part of this study is known as the Owlstone Medical Novel Insights (OMNI) method, and will be described in this section. Breath samples were collected using Owlstone Medical's ReCIVA[®] Breath Sampler. The ReCIVA Breath Sampler pre-concentrates breath samples onto adsorbent tubes, enabling a larger collection of air volume and offering the potential for greater sensitivity in detecting low-abundance compounds. Subjects breathed normally into the ReCIVA mouthpiece with a nose plug (Supplementary Fig. 1). Approximately 1.25 L of breath are collected onto each sorbent tube – with the analysis completed with 2.5 L: 2 tubes for analysis and 2 tubes for backup. This meant a total of 5 L of breath was collected per participant. This takes approximately 12–15 min of normal tidal breathing into the device to collect.

Ambient contamination was minimized using the CASPER[®] Portable Air Supply during breath sampling, which filters ambient air into the ReCIVA (Supplementary Fig. 1). The CASPER is a portable air supply that takes in room air, filters it to remove VOCs and particulates, and supplies it directly into the ReCIVA Breath Sampler for a subject to breathe into. The CASPER removes VOCs using a replaceable air filter pack that is filled with activated carbon; VOCs from the ambient air are adsorbed to the surface of the carbon. These tools combined were used to collect VOCs produced from the subject and eliminate VOCs that are re-breathed directly from the air in the room. Equal volumes of matched system background samples were collected immediately before each breath sample. Using internal, fast response pressure sensors, the ReCIVA Breath Sampler can monitor patient breathing patterns in real-time (see Supplementary Fig. 1). Alongside the paired software, these sensors estimate when the end-tidal fraction of breath is being exhaled, and the sampling pumps are automatically turned on and off at the necessary time to collect that breath

fraction. For background samples, all collection hardware was configured as if to collect a breath sample, but with the mouthpiece opening sealed and the software configured to sample continuously (no selection for specific fractions of breath). The exact settings are detailed in Supplementary Table 1.

Breath and background samples were analyzed using the Breath Biopsy OMNI settings (see Supplementary Materials for more details). The tubes were purged with a TD-100 (Markes International Ltd. Llantrisant, UK) and stored at a temperature of 4–8 °C for no more than 27 days before analysis. Breath and their paired background samples were liquid injected with a mix of eight deuterated internal standard compounds (Supplementary Table 2) solubilized in methanol and analyzed in the same sequence using TD-GC-MS. A series of straight-chain alkanes (C5–C16) were spiked onto a separate tube (50ng per alkane) and analyzed within each analytical sequence, to enable the calculation of retention indices. Analysis was conducted on the TD (Markes) – Q exactive Orbitrap (Thermo Fisher Scientific) high-resolution accurate mass spectrometry platform, utilizing the settings specified in Supplementary Table 3.

ReCIVA and CASPER breath collection, paired background sampling, TD-GC-MS analysis, and the feature extraction method are collectively known as the 'OMNI' method.

2.3 Feature extraction and data normalization

The resulting breath and background chromatograms (an example of a breath chromatogram is shown in Supplementary Fig. 2) were batch processed (spectral deconvolution, feature group clustering, and library matching to NIST17) utilizing the OMNI untargeted feature extraction method in Compound Discoverer (ver. 3.2, Thermo Scientific[™]), detailed in Supplementary Table 4. After feature extraction, all features were normalized using the measured peak area intensity response of spiked internal standard (IS) compounds to reduce analytical variability associated with TD-GC-MS. A hybrid correlation-retention time normalization

method was applied, where the Pearson correlation coefficients between each feature in each sample and every IS compound were calculated. Features that had a correlation coefficient ≥ 0.8 with any IS compound were normalized using that IS compound's response. When correlations for a given feature were below 0.8 for all ISs, the mean peak area response of the three IS compounds with the closest retention times to the feature was used for normalization.

2.4 Calculations comparing breath and paired system background samples

The three metrics were used to compare VOCs in breath samples and paired system background samples:

1. The standard deviation (SD) metric: A VOC was considered on-breath if the signal exceeded the mean of the system background signal plus 3 SDs in at least 50% of the breath samples from the cohort. A VOC is automatically on-breath if values are observed in less than 4 system backgrounds. Additionally, a feature observed in breath samples was automatically regarded "on-breath" if its signal was observed in fewer than 4 system background samples only.
2. The paired T-test approach metric: A VOC was regarded as on-breath if the paired breath/ system background samples were associated with a fold difference ≥ 2 and paired t-test one-tailed p -value ≤ 0.05 .
3. The Receiver Operating characteristic area under the curve [ROC-AUC] metric: A VOC was considered on-breath if the fold difference between breath and background was > 1 , and the calculated ROC-AUC value was ≥ 0.8 .

Each on-breath metric queries the VOC signal detected in breath samples, with respect to the system background, in different ways. This increases the confidence of a VOC's assignment as "on-breath" if it is calculated as such, by multiple metrics.

2.5 VOC identification using chemical standards

The candidate identities of on-breath VOCs were determined by matching the breath data against the NIST library and cross-checked against the human metabolome database (HMDB) (Westhoff et al., 2022). All NIST matches with a similarity index (SI) Match Factor ≥ 500 were considered for confirmation using purified standards. All NIST matches with a similarity index (SI) Match Factor of 500 or higher were considered for confirmation using purified standards. This threshold was chosen to reduce the risk of missing true matches due to spectral differences between the NIST

spectra and the experimental data. These differences can arise from variations in the analytical methods, instruments, and deconvolution parameters used to generate the spectra. Specifically, the in-house library data was generated using Orbitrap-MS, whereas the NIST spectra are based on quadrupole mass analyzer data. The VOCs with the highest SI score and present in the HMDB were prioritized. Additionally, a list of commonly reported VOCs with hypothesized biological relevance were compiled from a literature search and added to the candidate list.

A certified reference standard (minimum 95% purity) was sourced for each candidate compound and analyzed to generate spectra for matching against on-breath VOCs. Reference standards were dissolved in methanol, due to its suitability for GC-MS analysis in terms of expansion coefficient and solubility for each candidate compound. Two to sixteen standards were grouped into each mix using NIST17-reported retention index values to minimize co-elution risk. The prepared chemical mixes were then liquid injected onto Tenax TA-Carbograph-5TD sorbent tubes resulting in 50 ng on-tube mass per chemical and analyzed using the OMNI analytical method (see Supplementary) alongside a C5 to C16 straight chain alkane RI ladder.

Spectra for individual reference standards were identified by deconvolution (using Thermo Scientific GC Deconvolution plugin, using the peak detection settings in Supplementary Table 3) followed by cross-referencing with the NIST library (NIST 17 mainlib and replib). A background tube loaded with methanol was examined to ensure the peaks of interest were not derived from contamination during analytical processing. Mass spectral cleaning was performed to retain only the high-resolution accurate mass fragments suspected to derive from the reference standards. The final spectrum of each reference standard confirmed the on-breath VOC identities. Confirmation relied on the three breath chromatograms with the highest normalized peak area intensity demonstrating a successful match (forward and reverse similarity index (SI & RSI) above 800, retention index within +/- 2 units). Undetected candidate standards were re-analyzed at higher concentrations (150 ng on-tube mass) to increase detection probability.

3 Results

3.1 Distinguishing on-breath VOCs from background

Following the analysis of the 90 adult breath samples and paired system backgrounds using the OMNI method, 1471 unique features were present in $\geq 80\%$ of breath samples. Three metrics, detailed in the Methods section, were applied

to identify the subset of VOCs present in the breath at levels significantly above those in the system background (henceforth referred to as “on-breath”).

Figure 1A shows the total number of on-breath VOCs that were calculated using each of the 3 metrics. There is a significant overlap in the subset of on-breath features classified using each on-breath calculation metric. A total of 585 VOCs were identified as on-breath using any metric, and, of these, the majority ($328/585 = 56\%$) were on-breath by all metrics. Metric 1 was the most stringent, with most ($328/346 = 95\%$) of the features identified on-breath using metric 1 also on-breath by the other two metrics.

Metric 1 includes a flexible cut-off for the frequency of a VOC’s appearance on-breath at levels 3 SDs above background (Fig. 1B). In this analysis, a 50% frequency threshold was applied, restricting the subset of on-breath VOCs to 346 out of the total 1471 (22.3%) features (Fig. 1B). This threshold was chosen to emphasize the VOCs that are on-breath in the majority of samples but could be adjusted to accommodate other analyses. For example, if a more stringent threshold was deemed appropriate, fewer VOCs could be considered as on-breath.

The three metrics were chosen to provide complimentary insights into the potential composition of on-breath VOCs. Equally, metrics of differing stringencies, when combined into a panel of metrics, may give higher confidence to an on-breath identification. For example, being on-breath in multiple metrics at once, or in metrics with lower odds of a false positive may provide higher confidence that a VOC is indeed on breath, while still ensuring that a wide range of potentially on-breath VOCs are still captured by at least one metric.

3.2 Identified VOCs: chemical characteristics

A total of 148 (25% of 585 VOCs on-breath by any metric) VOCs were able to be assigned identities based on comparisons to reference standards analyzed on the same analytical method in this dataset. A total of 825 purified chemical standards were run to achieve this; 37% of NIST matches with SI scores over 800 were found to be the true identity of the on-breath VOC. Factors impeding the identification of the remaining on-breath features include poor matches against the NIST library due to differences in analytical methodologies used (as discussed above), logistical considerations (such as lack of standard availability, safety considerations and/or prohibitive cost) and potential spectral issues during deconvolution (such as co-elution or splitting), whereby the resulting VOC spectra may not be an accurate representation of a true compound. Possible avenues of further work to overcome these limitations include considering custom synthesis of standards not readily available off the shelf, along with applying tailored spectral deconvolution settings for the peak-rich regions of the breath sample chromatograms.

Of the 148 identified VOCs, 102 are on-breath by all three metrics, a single identified VOC is on-breath by metric 1 alone, three identified VOCs are on-breath by metric 2 alone, and nine identified VOCs are on-breath by metric 3 alone (Fig. 1A). A substantial portion of the on-breath VOCs that have been assigned formal identifications ($29/148 = 19.6\%$) were not classified as on-breath by metric 1, but they were on-breath by both metric 2 and metric 3 (Fig. 1A). While the three metrics have substantial overlap in the VOCs they determine to be on-breath, each metric contributed unique entries to the final pool of identified on-breath VOCs and may be appropriate for different VOCs

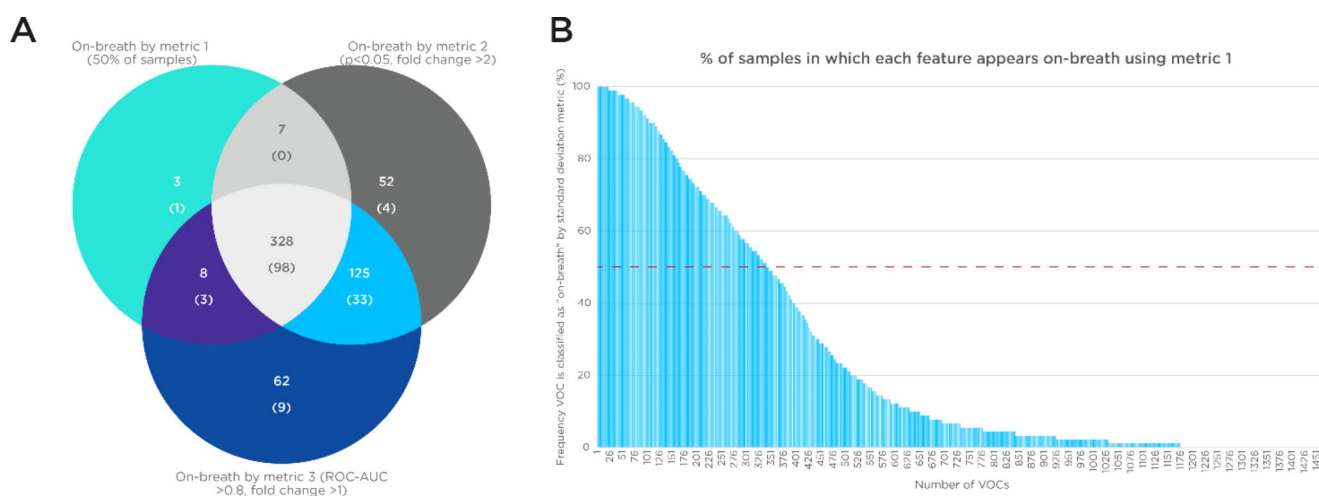


Fig. 1 A - Venn diagram showing the numbers of VOCs classified as on-breath by each metric, along with the number of those VOCs that have been identified, in brackets. B - Bar chart showing the frequency with which individual VOCs are classified as on-breath across all

samples using metric 1. The dotted line indicates the 50% threshold, restricting the number of on-breath VOCs above this cut-off to 346 of the 1471 total

or study designs. Additionally, to include VOCs that are on-breath in only a subset of the population, due to their uniqueness to a particular demographic, a separate analysis was carried out whereby features' on-breath status was calculated per collected demographic variable, by considering only the system background samples relevant to the specific sub-population. This resulted in 3 additional on-breath features: two unique to the age 70+ group (one of which was successfully identified as 1,3-Dimethylcyclohexane), and one unique to the 30+ BMI group.

A full list of the identities of the 148 on-breath VOCs are presented in Table 2.

4 Discussion

The challenges in studying breath VOCs are well-known in the research community. Healthy human breath profiles have been developed to understand how physiological conditions, including age, gender and circadian rhythms, can influence breath profiles (Sasiene et al., 2024). However, differentiation of on-breath compounds from background, and confirmation of their identities, remain major obstacles to advancing their applications in diagnostics and clinical settings. In this study, we present a list of 148 breath-associated (on-breath) chemically identified VOCs. The integrity of this data relies on stringent criteria for two key aspects: distinguishing VOCs from background contaminants and confirming their chemical identity. The on-breath VOCs presented in this study have been confidently chemically identified using MSI standards and were distinguishable from background contaminants through a robust methodology in a heterogeneous human population (spanning a range of ages, BMIs, and ethnicities). On-breath VOCs span 45 chemical classes, indicating that they comprise a diverse pool of chemical entities, and 62% have been previously reported in the literature in different biological matrices such as blood, urine, and fecal matter. However, the identified VOCs may also not necessarily be consistent with other studies in the literature due to differences in populations and analytical methodologies utilized previously.

It is also imperative to emphasize that on-breath VOCs can include both endogenously and exogenously generated VOCs (as exogenous VOCs can be very strongly breath-associated, especially those generated from internal sources such as the gut microbiome). Although not produced by the body, exogenous on-breath VOCs can interact widely with the host metabolome and demonstrate powerful utility in bridging the gap between research tools and breath-based clinical applications. One such example is indole, which is generated by the catabolism of tryptophan, mediated by the human gut microbiome; elevated levels of indole observed

in cirrhosis could be explained by impaired hepatic clearance, providing a plausible mechanistic relationship between a microbiome related VOC and a clinical diagnosis (Ferrandino et al., 2023). In terms of well-established clinical applications, the gold standard diagnostic test used in gut health clinics for small intestinal bacterial overgrowth (SIBO) involves the ingestion of an exogenous substrate (lactulose). If SIBO is present, this synthetic sugar is metabolized by bacteria in the small intestine to produce molecular hydrogen, detectable on breath. Similarly, limonene, a VOC generated by dietary exposure was observed to be elevated in the breath of subjects with cirrhosis compared to controls in multiple studies (Dadamio et al., 2012; Fernández del Río et al., 2015), suggesting that reduced liver function and impaired hepatic perfusion induce limonene accumulation in the body resulting in elevated levels in breath. These alterations make limonene a candidate biomarker for non-invasive cirrhosis detection using a breath test (Ferrandino et al., 2023).

Acetone, isoprene, and indole can be used as a starting point to assess the consistency of the results of this study with breath compositions reported elsewhere, as these are some of the most abundant and commonly identified breath VOCs (Drabińska et al., 2021). All three of these compounds were frequently found to be on-breath within this population, supporting the replicability of this study's results. Isoprene has been associated previously with a broad range of disease states, however, there are doubts over how useful of a biomarker breath isoprene currently is due to the lack of specificity to certain disease states, and sensitivity to individual breathing patterns and movement (Mochalski et al., 2023). Breath isoprene has recently been mechanistically associated with skeletal muscle metabolic activity using a multi-omic approach (Sukul et al., 2023; Mochalski et al., 2024), demonstrating endogenous origin. The majority of breath isoprene is produced through the IDI2 protein that is only present within skeletal-myocellular peroxisomes (Sukul et al., 2023), and therefore supports the observation that breath isoprene abundance increases after exercise (Chou et al., 2024; Pugliese et al., 2022). This understanding of the body's mechanistic origin has helped to associate isoprene with a specific physiological process and could help establish what clinically useful information could be gained by the use of breath isoprene as a biomarker.

As the current study only aimed to characterize the composition of normal human breath, the identified on-breath VOCs currently cannot suggest metabolic pathway changes, however, certain valuable insights can be gained. For example, acetic acid and propionic acid were both found to be on-breath and are two well-characterized short-chain fatty acids (SCFAs) associated with the gut microbiome. SCFAs are considered exogenous VOCs because they are produced

Table 2 The table shows all 148 on-breath VOCs that were able to be confidently chemically identified in this study, along with their InChI key identifier and chemical classification. The criteria they passed (TRUE) and failed (FALSE) to be classified as on-breath by the three metrics is also shown. The compounds are sorted by alphabetical order of chemical class

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
Bromodichloromethane	FMWLUWQPKEAR-UHFFFAOYSA-N	Alkyl halides	Halomethanes	41	FALSE	TRUE	TRUE
Chlorodibromomethane	GATVJKZLVQHOMN-UHFFFAOYSA-N	Alkyl halides	Halomethanes	59	TRUE	TRUE	TRUE
Allyl methyl disulfide	XNZOTQPMYCTBZ-UHFFFAOYSA-N	Allyl sulfur compounds		27	FALSE	TRUE	TRUE
Allyl methyl sulfide	NVLPQIPTCCLBEU-UHFFFAOYSA-N	Allyl sulfur compounds		96	TRUE	TRUE	TRUE
Imidazole	RAXXELZNTBOGNW-UHFFFAOYSA-N	Azoles	Imidazoles	80	TRUE	TRUE	TRUE
Thiazole	FZWLAABWBMGSTO-UHFFFAOYSA-N	Azoles	Thiazoles	96	TRUE	TRUE	TRUE
1,3-Dichlorobenzene	ZPQPVIELGIULI-UHFFFAOYSA-N	Benzene and substituted derivatives	Halobenzenes	32	FALSE	FALSE	TRUE
1,4-Dichlorobenzene	OCJBOOLMMGQPQU-UHFFFAOYSA-N	Benzene and substituted derivatives	Halobenzenes	24	FALSE	FALSE	TRUE
2,5-Dimethylstyrene	DBWWINQJTYDFK-UHFFFAOYSA-N	Benzene and substituted derivatives	Styrenes	51	FALSE	TRUE	TRUE
2,6-Diphenylphenol	ATGFTMUSEPNJD-UHFFFAOYSA-N	Benzene and substituted derivatives	Terphenyls	68	TRUE	TRUE	TRUE
2-Cyclohexylphenol	MVRPPTGLVPEMPI-UHFFFAOYSA-N	Benzene and substituted derivatives	Cyclohexylphenols	51	TRUE	TRUE	TRUE
2-Phenylethanol	WRMNCZEMHIOC-P-UHFFFAOYSA-N	Benzene and substituted derivatives		42	FALSE	TRUE	TRUE
Benzene	UHOVQNZJYSORNB-UHFFFAOYSA-N	Benzene and substituted derivatives		27	FALSE	TRUE	TRUE
Chlorobenzene	MVPPADPHJFYWMZ-UHFFFAOYSA-N	Benzene and substituted derivatives	Halobenzenes	44	FALSE	TRUE	TRUE
Cumene	RWGFKTVRMDUZSP-UHFFFAOYSA-N	Benzene and substituted derivatives	Cumenes	69	TRUE	TRUE	TRUE
Fluorobenzene	PYLWMHQQBFSUBP-UHFFFAOYSA-N	Benzene and substituted derivatives	Halobenzenes	33	FALSE	TRUE	TRUE
Isophthalaldehyde	IZALUMVGBVKPJ-D-UHFFFAOYSA-N	Benzene and substituted derivatives	Benzoyl derivatives	17	FALSE	FALSE	TRUE
m-Cymene	XCYJPXQACVEIOS-UHFFFAOYSA-N	Benzene and substituted derivatives	Cumenes	87	TRUE	TRUE	TRUE
Methyl salicylate	OSWPMRLSEDDHDF-UHFFFAOYSA-N	Benzene and substituted derivatives	Benzoic acids and derivatives	20	FALSE	TRUE	TRUE
Phenylacetaldehyde	DTUQGWVMVHBKE-UHFFFAOYSA-N	Benzene and substituted derivatives	Phenylacetaldehydes	18	FALSE	FALSE	TRUE
Phenylhydroquinone	XCZKKZXWBOGPA-UHFFFAOYSA-N	Benzene and substituted derivatives	Biphenyls and derivatives	46	FALSE	TRUE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
Toluene	YXFVVABEGXRONW-UHFFFAOYSA-N	Benzene and substituted derivatives	Toluenes	12	FALSE	FALSE	TRUE
2-Coumaranone	ACZGCSMSTYWDQ-UHFFFAOYSA-N	Benzofurans		50	TRUE	TRUE	TRUE
2-Methylbenzofuran	GBGPVUAOTCNZPT-UHFFFAOYSA-N	Benzofurans		21	FALSE	TRUE	TRUE
Chromone	OTAFHZMPRISVEM-UHFFFAOYSA-N	Benzopyrans	1-benzopyrans	60	TRUE	FALSE	TRUE
Galaxolide	ONKNPOPIGWHAQC-UHFFFAOYSA-N	Benzopyrans	2-benzopyrans	53	TRUE	TRUE	TRUE
Formamide	ZHNUHDYFZUAESO-UHFFFAOYSA-N	Carboximide acids and derivatives	Carboximide acids	97	TRUE	TRUE	TRUE
Acetic acid	QTBSBXVTEAMEQO-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acids	70	TRUE	TRUE	TRUE
Methyl acetate	KXKVLQRXCPHEJC-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acid derivatives	98	TRUE	TRUE	TRUE
Methyl acrylate	BAPJBEWLBFYGME-UHFFFAOYSA-N	Carboxylic acids and derivatives	Acrylic acids and derivatives	96	TRUE	TRUE	TRUE
Methyl methacrylate	VVQNEPGJFQJSBK-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acid derivatives	3	FALSE	TRUE	TRUE
N, N-Dimethylformamide	ZMXDDKWL CZADIW-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acid derivatives	20	FALSE	TRUE	TRUE
N-Methylsuccinimide	KYEACNNYFNZCST-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acid derivatives	83	TRUE	TRUE	TRUE
Propionamide	QLNJFJADRCOGBJ-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acid derivatives	67	TRUE	TRUE	TRUE
Propionic acid	XBDQKXXYIPTUBI-UHFFFAOYSA-N	Carboxylic acids and derivatives	Carboxylic acids	62	TRUE	TRUE	TRUE
Pyrazine	KYQCOXFCLRLTKLS-UHFFFAOYSA-N	Diazines	Pyrazines	77	TRUE	TRUE	TRUE
2,3-Dihydrofuran	JKTCBAGSMQJFNL-UHFFFAOYSA-N	Dihydrofurans		58	TRUE	TRUE	TRUE
4-Methyl-2(5H)-furanone	ZZEYQBQZKUKWKY-UHFFFAOYSA-N	Dihydrofurans	Furanones	84	TRUE	TRUE	TRUE
1,4-Dioxane	RYHBNJHYFVUHQT-UHFFFAOYSA-N	Dioxanes	1,4-dioxanes	90	TRUE	TRUE	TRUE
1,4-Dioxene	HIZVCIORGCREW-UHFFFAOYSA-N	Dioxins	Para dioxins	44	FALSE	TRUE	TRUE
2-Methyl-1,3-dioxolane	HTWIZMNMWYQRN-UHFFFAOYSA-N	Dioxolanes	1,3-dioxolanes	79	TRUE	TRUE	TRUE
1-Hexadecanol	BXWNKGSJHAJOGX-UHFFFAOYSA-N	Fatty Acyls	Fatty alcohols	56	TRUE	TRUE	TRUE
Isopropyl myristate	AXISY YRBXTVTFY-UHFFFAOYSA-N	Fatty Acyls	Fatty acid esters	19	FALSE	FALSE	TRUE
Isopropyl palmitate	XUGNVMKQXJXZCD-UHFFFAOYSA-N	Fatty Acyls	Fatty acid esters	61	TRUE	TRUE	TRUE
Methyl furan-3-carboxylate	ZKHQSQYLKSSYIP-UHFFFAOYSA-N	Furans	Furoic acid and derivatives	90	TRUE	TRUE	TRUE
2,3,5-Trimethylfuran	NIXZFRUNHWKHEC-UHFFFAOYSA-N	Heteroaromatic compounds		98	TRUE	TRUE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
2,5-Dimethylfuran	GSNUFIFRDBKVE-UHFFFAOYSA-N	Heteroaromatic compounds		96	TRUE	TRUE	TRUE
2-Furonitrile	YXDXXGXWFWJCXEB-UHFFFAOYSA-N	Heteroaromatic compounds		96	TRUE	TRUE	TRUE
2-Methylfuran	VQKFNUFAXTZWDK-UHFFFAOYSA-N	Heteroaromatic compounds		98	TRUE	TRUE	TRUE
3-Methylthiophene	QENGPZGAWFQWCZ-UHFFFAOYSA-N	Heteroaromatic compounds		98	TRUE	TRUE	TRUE
Pyrrrole	KAESVJOAVNADME-UHFFFAOYSA-N	Heteroaromatic compounds		92	TRUE	TRUE	TRUE
3-Methylindole	ZFRKQXVRDFCRJG-UHFFFAOYSA-N	Indoles and derivatives	Indoles	93	TRUE	TRUE	TRUE
Indole	SIKJAQRHWYJAI-UHFFFAOYSA-N	Indoles and derivatives	Indoles	97	TRUE	TRUE	TRUE
gamma-Valerolactone	GAEKPEKOKCEMS-UHFFFAOYSA-N	Lactones	Gamma butyrolactones	57	TRUE	TRUE	TRUE
1,3-Dioxol-2-one	VAYTZRYEBVHVLE-UHFFFAOYSA-N	Organic carbonic acids and derivatives	Carbonic acid diesters	66	TRUE	TRUE	TRUE
Dimethyl disulfide	WQOXQRCZOLPYPM-UHFFFAOYSA-N	Organic disulfides	Dialkyl disulfides	98	TRUE	TRUE	TRUE
Dimethyl trisulfide	YWHLKYYPLRWGE-UHFFFAOYSA-N	Organic trisulfides		92	TRUE	TRUE	TRUE
1,1,1-Trichloroethane	UOCLXMDMGBRAIB-UHFFFAOYSA-N	Organochlorides		14	FALSE	TRUE	TRUE
Dodecamethylcyclohexasiloxane	IUMSDRXLFWAGNT-UHFFFAOYSA-N	Organometalloid compounds	Organosilicon compounds	29	FALSE	TRUE	TRUE
Hexadecamethylheptasiloxane	NFVSFLUJRHRJSJG-UHFFFAOYSA-N	Organometalloid compounds	Organosilicon compounds	50	TRUE	TRUE	TRUE
Acetonitrile	WEVYAHXRMPXWCK-UHFFFAOYSA-N	Organonitrogen compounds	Organic cyanides	16	FALSE	TRUE	TRUE
Trimethylamine	GETQZCLCWQTVFV-UHFFFAOYSA-N	Organonitrogen compounds	Amines	91	TRUE	TRUE	TRUE
1-Propanol	BDERNNEJNPAEC-UHFFFAOYSA-N	Organooxygen compounds	Alcohols and polyols	57	TRUE	TRUE	TRUE
2-(2-Methoxyethoxy)ethanol	SBASXUCJHJRPEV-UHFFFAOYSA-N	Organooxygen compounds	Ethers	59	TRUE	TRUE	TRUE
2,3-Butanediol	OWBTYPJTUOEWEK-UHFFFAOYSA-N	Organooxygen compounds	Alcohols and polyols	73	TRUE	TRUE	TRUE
2,3-Butanedione	QSIXEFYDPANLFS-UHFFFAOYSA-N	Organooxygen compounds	Carbonyl compounds	74	TRUE	TRUE	#REF!
2-Acetyl-5-methylfuran	KEFJLCGVTHRGAH-UHFFFAOYSA-N	Organooxygen compounds	Carbonyl compounds	93	TRUE	TRUE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
2-Acetyl furan	IEMMBWWQXVXBEU-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	96	TRUE	TRUE	TRUE
2-Acetylpyridine	AJKVQEKCUACUMD-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	68	TRUE	TRUE	TRUE
2'-Hydroxyacetophenone	JECYUBVRTQDVAT-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	87	TRUE	TRUE	TRUE
2-Methoxyethanol	XNWRZJHXBZDAG-UHFFFAOYSA-N	Organoxygen compounds	Ethers	62	TRUE	TRUE	TRUE
2-Methyl-2-butenal	ACWQBUSCFJPUN-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	98	TRUE	TRUE	TRUE
2-Methyl-3-buten-2-ol	HNVRRHSLBLFLIG-UHFFFAOYSA-N	Organoxygen compounds	Alcohols and polyols	91	TRUE	TRUE	TRUE
2-Pentanone	XNLICIUVMPIYHGG-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	96	TRUE	TRUE	TRUE
2-Propenol	XXROGKLTUQVRX-UHFFFAOYSA-N	Organoxygen compounds	Alcohols and polyols	32	FALSE	TRUE	TRUE
3-Buten-2-one	FUSUHKVFWTUUBE-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	99	TRUE	TRUE	TRUE
3-Thiophenecarboxaldehyde	RBIGKSZIQCTIJF-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	96	TRUE	TRUE	TRUE
4-Heptanone	HCFAJYNVAYBARA-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	92	TRUE	TRUE	TRUE
5-Methylfurfural	OUDFNZMQXZILJD-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	84	TRUE	TRUE	TRUE
Acetaldehyde	IKHGUXGNUTLKF-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	47	FALSE	TRUE	TRUE
Acetone	CSCPPACGZOOCGX-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	88	TRUE	TRUE	TRUE
Benzofuran-2-yl(phenyl)methanone	DZRINLPOTUVETG-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	26	FALSE	TRUE	TRUE
Diethylene glycol monoethyl ether	XXJWXESWEXIICW-UHFFFAOYSA-N	Organoxygen compounds	Ethers	61	TRUE	TRUE	TRUE
Ethylene glycol	LYCAIKOWRPUZTN-UHFFFAOYSA-N	Organoxygen compounds	Alcohols and polyols	47	FALSE	TRUE	TRUE
Furan-2,5-dicarbonyl aldehyde	PXJJKVNIMAZHCB-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	17	FALSE	FALSE	TRUE
Furfural	HYBBIBJNHNGZAN-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	19	FALSE	FALSE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
Methacrolein	STNJBCKSHOAVAJ-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	99	TRUE	TRUE	TRUE
Phenacyl acetate	BGAXCPSNMHVHJC-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	50	TRUE	TRUE	TRUE
Propylene glycol	DNIAPMSPWPWGF-UHFFFAOYSA-N	Organoxygen compounds	Alcohols and polyols	33	FALSE	TRUE	FALSE
Pyrrrole-2-carboxaldehyde	ZSKGQVFTSEPTJ-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	80	TRUE	TRUE	TRUE
Salicylaldehyde	SMQUZDBALVYZAC-UHFFFAOYSA-N	Organoxygen compounds	Carbonyl compounds	73	TRUE	TRUE	TRUE
Triethylene glycol monomethyl ether	JLGLQAWTXXGVEM-UHFFFAOYSA-N	Organoxygen compounds	Ethers	47	FALSE	TRUE	FALSE
Carbon disulfide	QGJOPFRUIJSHPO-UHFFFAOYSA-N	Other non-metal organides	Other non-metal sulfides	73	TRUE	TRUE	TRUE
Eucalyptol	WEEGYLXZBRQIMU-UHFFFAOYSA-N	Oxanes	Oxanes	94	TRUE	TRUE	TRUE
1,6-Anhydro-beta-D-galactose	TWNIBLMWSKIRAT-UHFFFAOYSA-N	Oxepanes	Oxepanes	32	FALSE	TRUE	FALSE
Phenalen-1-one	WWBGWPHHLRSTFI-UHFFFAOYSA-N	Phenalenes	Phenalenones	34	FALSE	TRUE	TRUE
2-(3-Methylphenoxy)ethanol	FDBXUXVQIOQYIX-UHFFFAOYSA-N	Phenol ethers	Phenol ethers	77	TRUE	TRUE	TRUE
Anethole	RUVINXPYWBJD-UHFFFAOYSA-N	Phenol ethers	Anisoles	86	TRUE	TRUE	TRUE
4-Ethylphenol	HXDZKJGKXYMEW-UHFFFAOYSA-N	Phenols	1-hydroxy-2-unsubstituted benzenoids	76	TRUE	TRUE	TRUE
Guaiacol	LHGVFZTFXWLCU-UHFFFAOYSA-N	Phenols	Methoxyphenols	67	TRUE	TRUE	TRUE
Hydroquinone	QIGBRXMKCJVMJ-UHFFFAOYSA-N	Phenols	Benzenediols	39	FALSE	TRUE	TRUE
o-Cresol	QWVGKYWNOKOFNN-UHFFFAOYSA-N	Phenols	Cresols	71	TRUE	TRUE	TRUE
Phenol	ISWSDIOOJBQZ-UHFFFAOYSA-N	Phenols	1-hydroxy-4-unsubstituted benzenoids	97	TRUE	TRUE	TRUE
1,4-Cineole	RFFOTVCVTJUTAD-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	28	FALSE	TRUE	TRUE
3-Carene	BQOFWKZOCNGFEC-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	82	TRUE	TRUE	TRUE
alpha-Curcumene	VMYXUZSZMNBRCN-UHFFFAOYSA-N	Preneol lipids	Sesquiterpenoids	21	FALSE	TRUE	FALSE
alpha-Pinene	GRWFGVWFZKLTU-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	96	TRUE	TRUE	TRUE
alpha-Thujene	KQAZVFVQERWLN-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	63	TRUE	TRUE	TRUE
beta-Phellandrene	LFJQCDVYDGGFCH-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	83	TRUE	TRUE	TRUE
beta-Pinene	WTARULDDTDQWMU-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	93	TRUE	TRUE	TRUE
Caryophyllene	NPNUFJAVOONJIE-UHFFFAOYSA-N	Preneol lipids	Sesquiterpenoids	60	FALSE	TRUE	TRUE
Limonene	XMGQYMWWDQXJHM-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	87	TRUE	TRUE	TRUE
Menthofuran	YGWXXYGDYFJU-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	76	TRUE	TRUE	TRUE
Ocimene	IHPKGUQCSINRJ-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	60	TRUE	TRUE	TRUE
p-Menthan-3-one	NELGAXVYCFJBMK-UHFFFAOYSA-N	Preneol lipids	Monoterpenoids	73	TRUE	TRUE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
Valencene	QEBNVLSCGVZOH-UHFFFAOYSA-N	Preneol lipids	Sesquiterpenoids	47	FALSE	TRUE	TRUE
2-Methylpyridine	BSKHPKMHHTQYZBB-UHFFFAOYSA-N	Pyridines and derivatives	Methylpyridines	94	TRUE	TRUE	TRUE
3-Methylpyridine	ITQTTZVARXURQS-UHFFFAOYSA-N	Pyridines and derivatives	Methylpyridines	90	TRUE	TRUE	TRUE
3-Pyridinecarboxaldehyde	QJZUKDFHGGYHMC-UHFFFAOYSA-N	Pyridines and derivatives	Pyridine carboxaldehydes	83	TRUE	TRUE	TRUE
Pyridine	JUJWROOIHZHMGM-UHFFFAOYSA-N	Pyridines and derivatives	Pyridines and derivatives	88	TRUE	TRUE	TRUE
1-Methylpyrrole	OXHNLMTVIGZXS-G-UHFFFAOYSA-N	Pyrroles	Substituted pyrroles	54	TRUE	TRUE	TRUE
2-Methyl-1 H-pyrrole	TVCXVUHHCUYLGX-UHFFFAOYSA-N	Pyrroles	Substituted pyrroles	23	FALSE	TRUE	TRUE
1,1,3-Trimethylcyclohexane	PYOLJOJPCRPD-UHFFFAOYSA-N	Saturated hydrocarbons	Cycloalkanes	16	FALSE	TRUE	TRUE
Heptane	IMNFDUFMRHMDMM-UHFFFAOYSA-N	Saturated hydrocarbons	Alkanes	12	FALSE	TRUE	TRUE
Nonane	BKIMMITUMNQMO-UHFFFAOYSA-N	Saturated hydrocarbons	Alkanes	17	FALSE	TRUE	TRUE
Octane	TVMXXDCGIABBOFY-UHFFFAOYSA-N	Saturated hydrocarbons	Alkanes	49	FALSE	TRUE	TRUE
Diphenylmaleic anhydride	OUCFCNZIUTYBH-UHFFFAOYSA-N	Stilbenes		29	FALSE	TRUE	TRUE
Dimethyl sulfoxide	IAZDPXIOMUYVVGZ-UHFFFAOYSA-N	Sulfoxides		80	TRUE	FALSE	TRUE
Tetrahydrofuran	WYURNTSHIVDZCO-UHFFFAOYSA-N	Tetrahydrofurans		28	FALSE	FALSE	TRUE
Bis(methylthio)methane	LOCDPORVFOGCR-UHFFFAOYSA-N	Thioacetals	Dithioacetals	93	TRUE	FALSE	FALSE
Methyl thiocyanate	VYHVQEYOFIYNJP-UHFFFAOYSA-N	Thiocyanates		98	TRUE	TRUE	TRUE
Dimethyl sulfide	QMMFVYPAHWCMMS-UHFFFAOYSA-N	Thioethers	Dialkylthioethers	97	TRUE	TRUE	TRUE
Methyl propyl sulfide	ZOASGOXWEHUTKZ-UHFFFAOYSA-N	Thioethers	Dialkylthioethers	96	TRUE	TRUE	TRUE
Methenamine	VKYKSIONXXAKP-UHFFFAOYSA-N	Triazinanes	1,3,5-triazinanes	93	TRUE	TRUE	TRUE
1,3,5-Triazine	JIHQDMXYFUGFV-UHFFFAOYSA-N	Triazines	1,3,5-triazines	90	TRUE	TRUE	TRUE
1-Butene	VXNZUUAINFGPBY-UHFFFAOYSA-N	Unsaturated hydrocarbons	Unsaturated aliphatic hydrocarbons	72	TRUE	TRUE	TRUE
1-Docosene	SPURMHFLEKVAAS-UHFFFAOYSA-N	Unsaturated hydrocarbons	Unsaturated aliphatic hydrocarbons	48	FALSE	TRUE	TRUE
1-Methylcyclopentene	ATQFUXWBVZUTKO-UHFFFAOYSA-N	Unsaturated hydrocarbons	Branched unsaturated hydrocarbons	56	TRUE	FALSE	TRUE
2-Pentene	QMMOXUPEWRXHS-UHFFFAOYSA-N	Unsaturated hydrocarbons	Unsaturated aliphatic hydrocarbons	32	FALSE	TRUE	TRUE
3-Methyl-1-butene	YHQXBTXEYZIYOV-UHFFFAOYSA-N	Unsaturated hydrocarbons	Unsaturated aliphatic hydrocarbons	59	TRUE	TRUE	TRUE

Table 2 (continued)

On-breath VOC ID	InChI key	Class	Subclass	% of breath samples on-breath in by metric 1	On-breath status		
					Metric 1 (mean + std dev)	Metric 2 (paired t-test)	Metric 3 (ROC-AUC)
3-Methyl-2-pentene	BEQRRJLJVQAQ-UHFFFAOYSA-N	Unsaturated hydrocarbons	Branched unsaturated hydrocarbons	74	TRUE	TRUE	TRUE
Cyclohexene	HGCIXCUEYOPUTN-UHFFFAOYSA-N	Unsaturated hydrocarbons	Olefins	56	TRUE	TRUE	TRUE
Cyclopentene	LPIQUOYDBNQMRZ-UHFFFAOYSA-N	Unsaturated hydrocarbons	Olefins	90	TRUE	TRUE	TRUE
gamma-Terpinene	YKFLAYDHMOASTY-UHFFFAOYSA-N	Unsaturated hydrocarbons	Branched unsaturated hydrocarbons	93	TRUE	TRUE	TRUE
Isoprene	RRHGJUQNOFWUDK-UHFFFAOYSA-N	Unsaturated hydrocarbons	Branched unsaturated hydrocarbons	99	TRUE	TRUE	TRUE
Isopropenylacetylene	BOFLDKIFLFLJA-UHFFFAOYSA-N	Unsaturated hydrocarbons	Branched unsaturated hydrocarbons	88	TRUE	TRUE	TRUE
Penta-1,3-diene	PMJHHCWVYXUKFD-UHFFFAOYSA-N	Unsaturated hydrocarbons	Olefins	79	TRUE	TRUE	TRUE
Tetrachloroethylene	CYTYCFOTNPOANT-UHFFFAOYSA-N	Vinyl halides	Vinyl chlorides	71	TRUE	TRUE	TRUE

by microbial fermentation of dietary fiber and are thought to diffuse into local blood vessels of the gastrointestinal tract, travel via the blood, and enter the breath through alveolar exchange. The microbially-formed gas hydrogen produced in the gastrointestinal tract is rapidly detectable in the breath through this mechanism, and is therefore currently used in the clinic to diagnose conditions such as small intestinal bacterial overgrowth (Pitcher et al., 2022; Read et al., 1985; Sachdev & Pimentel, 2013). The abundance level of SCFAs has been implicated in multiple health contexts, including cancer, neurodegenerative disease, and inflammatory bowel disease (Duiser & de Zoete, 2023; Majumdar et al., 2023; Ney et al., n.d.; Parada Venegas et al., 2019; van Vorstenbosch et al., 2023; Wang et al., 2023). In addition, their multiple signaling roles have become increasingly appreciated for their potential impacts on human health (Louis & Flint, 2017; Miller & Wolin, 1996). Therefore, the SCFAs could serve as breath biomarkers of disease much like hydrogen and methane breath tests in the future. Their characterization on-breath and the development of reference ranges in a healthy population is essential for this development, of which this study provides a useful starting point.

Isoprene, acetic and propionic acid are examples of the connections between the identified on-breath compounds and the literature, but there have been many more associations with a broader range of these compounds with physiological processes in the literature, such as carbon disulfide, dimethyl sulfide, and dimethyl disulfide (Carrion et al., 2015; Di Cagno et al., 2011; Grabowska-Polanowska et al., 2017; Preter et al., 2015). While the mechanism behind the appearance of certain on-breath VOCs in this dataset remain unknown, crucially, their identities have been verified. This suggests that the discovery of novel pathways and a new understanding of physiological processes can occur if the levels of these VOCs are found to change in disease cohorts.

In this study, the system background samples were collected from the entire equipment flow path of air to capture all possible sources of VOC background in the sampling process. Moisture levels may differ between background samples and breath. Given that humidity and temperature are key variables for VOC capture, mitigations have been implemented to reduce this differential: hydrophobic sorbents are used in our TD tubes and all samples are dry purged prior to analysis. We acknowledge that this minimizes the carryover risk, but not the instantaneous risk. Despite our best efforts, breath and blank samples are ultimately different sample types and there may be small analytical differences introduced as a result. We did not consider this to have a major impact to the conclusions we draw from this article.

The analytical methodology utilized offers high mass accuracy and resolution capability, resulting in precise mass measurement of ions and separation of closely spaced

mass peaks. These functionalities provided high specificity by accurately determining molecular formulas and distinguishing between different VOCs with similar masses, respectively, thereby increasing the accuracy of VOC identification. For background correction, this study built on the alveolar gradient approach and utilized three metrics to compare breath and background signals, which can account for the variability of VOCs in the samples while also being practical for field implementation. Stringent processes were applied to the identification of VOCs using reference standards run on the same analytical method and instrument, accurate mass (within 5 ppm tolerance) to compare fragmentation spectra, and an alkane ladder to enable a strict RI match. VOCs were only identified if they could be characterized as “on-breath” by pre-defined metrics used to compare breath to representative system background samples. These metrics are intended to capture the widest range of on-breath VOCs, while providing sufficient stringency. In the future, they can be used alone or in combination, depending on the study design and VOCs of interest. The multiple metrics not only expand our knowledge of common on-breath VOCs but also allow differentiation from background, increasing confidence when comparing VOCs across different study cohorts and methodologies.

Despite the stringent methodology used to exclude background contaminants, it is still likely that certain on-breath compounds arise from processes unrelated to underlying physiology. It is also possible that some breath VOCs may be currently below our effective detection limit due to the presence and variation of VOCs in the background, and systematically reducing background contamination could improve this. The list of on-breath VOCs provides a starting point for this work, as targeted analyses can help identify where background contamination may be reduced and which specific chemicals, or groups of chemicals are likely to be affected. Future work will expand the capabilities of detection via sampling onto different sorbent beds and using different GC-column chemistries, broadening the range of VOCs covered. This study included a heterogeneous cohort, enabling a comprehensive characterization of breath composition. It is likely that VOCs will be significantly elevated in diseased cohorts such that they may be characterized as on-breath in diseased populations alone. Therefore, future efforts should focus on identifying and quantifying on-breath VOCs in both heterogeneous and diseased populations to build a platform for cross-population comparison.

The on-breath VOCs identified in this study can offer utility to the breath field in numerous ways: they can serve as targets for optimizing breath measurement platforms, enabling efficient and accurate identification of potential biomarkers. With appropriate controls, these VOCs may also facilitate comparison of data across studies for cross-validation of

results. Furthermore, the list of VOCs can be utilized to optimize precise and accurate measurement of breath via informed selection of patient preparation, breath collection, sample storage, and sample analysis. This optimization process will preserve biological variability while minimizing technical variability, thereby advancing the reliability and reproducibility of breath-based biomarker research. This list of chemically confirmed on-breath VOCs distinguished from background contaminants lays the foundation for the development of the Breath Biopsy VOC Atlas[®], an ongoing project to develop a database of chemically identified breath VOCs complete with on-breath status, and quantified reference ranges across different cohorts, including different disease states. Biological interpretation of VOCs in the breath will significantly help to confidently assign on-breath VOC status, and therefore adding mechanistic understanding of breath VOCs in the literature is important future work. Work is ongoing to build the VOC Atlas as a reference database on confirmed compounds in the breath alongside their scientific context in the literature, utilizing the robust OMNI method and the on-breath VOC list presented in this work.

5 Conclusion

Through the development of a robust methodology, this study collected and compared breath and background samples of a heterogeneous human population to identify on-breath VOCs. These on-breath VOCs can serve as a reference for breath researchers to improve confidence that their results are capturing truly on-breath VOCs, and it will continue to expand as additional VOCs are identified using reference standards. Future work will expand the list to include a broad range of populations and physiologies to capture the diversity of on-breath VOCs. By continuing to compare background samples collected and analyzed in the same manner as their breath samples, VOCs confidently identified as being on-breath can be the basis for future biomarker investigations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11306-024-02163-6>.

Author contributions J.G, F.vS, and W.A designed the study. W.A, S.C, A.R, E.L, E.M, M.S, A.H, S.L, C.CP, L.NS, and S.S performed the research. W.A, S.C, and O.B analyzed the data. W.A, H.C, M.B, J.G, and A.H wrote the paper. J.G, W.A, M.B, H.C, P.G, B.B, and M.A critically reviewed and revised the paper.

Declarations

Ethical statement All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Hel-

sinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. The study supporting this work was approved by the Reading Independent Ethics Committee RIEC: 290620-1.

Competing interests BB is the CEO and Co-Founder of Owlstone Medical Ltd. WA, HC, JG, MB, OB, SC, PG, AR, EL, AH, MS, SL, EM, CCP, LND, SS, and MA are, or were employees of Owlstone Medical Ltd at the time this work was completed.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Babu, M., & Snyder, M. (2023). Multi-omics profiling for health. *Molecular & Cellular Proteomics*, 22(6), 100561. <https://doi.org/10.1016/j.mcpro.2023.100561>
- Carrión, O., Curson, A. R. J., Kumaresan, D., Fu, Y., Lang, A. S., Mercadé, E., & Todd, J. D. (2015). A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nature Communications*, 6, 6579. <https://doi.org/10.1038/ncomms7579>
- Chou, H., Arthur, K., Shaw, E., Schaber, C., Boyle, B., Allsworth, M., Kelley, E. F., Stewart, G. M., Wheatley, C. M., Schwartz, J., Fermoy, C. C., Ziegler, B. L., Johnson, K. A., Robach, P., Basset, P., & Johnson, B. D. (2024). Metabolic insights at the finish line: Deciphering physiological changes in ultramarathon runners through breath VOC analysis. *Journal of Breath Research*, 18(2). <https://doi.org/10.1088/1752-7163/ad23f5>
- Costello, B. L., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., Osborne, D., & Ratcliffe, N. M. (2014). A review of the volatiles from the healthy human body. *Journal of Breath Research*, 8(1), 014001. <https://doi.org/10.1088/1752-7155/8/1/014001>
- Dadamio, J., Van den Velde, S., Laleman, W., Van Hee, P., Coucke, W., Nevens, F., & Quirynen, M. (2012). Breath biomarkers of liver cirrhosis. *Journal of Chromatography B Analytical Technologies in the Biomedical and Life Sciences*, 905, 17–22. <https://doi.org/10.1016/j.jchromb.2012.07.025>
- del Fernández, R., O'Hara, M. E., Holt, A., Pemberton, P., Shah, T., Whitehouse, T., & Mayhew, C. A. (2015). Volatile biomarkers in breath associated with liver cirrhosis—comparisons of pre- and post-liver transplant breath samples. *EBioMedicine*, 2(9), 1243–1250. <https://doi.org/10.1016/j.ebiom.2015.07.027>
- Di Cagno, R., De Angelis, M., De Pasquale, I., Ndagijimana, M., Vernocchi, P., Ricciuti, P., Gagliardi, F., Laghi, L., Creccchio, C., Guerzoni, M. E., Gobetti, M., & Francavilla, R. (2011). Duodenal and faecal microbiota of celiac children: Molecular, phenotype and metabolome characterization. *BMC Microbiology*, 11(1), 219. <https://doi.org/10.1186/1471-2180-11-219>
- Di Gilio, A., Palmisani, J., Ventrella, G., Facchini, L., Catino, A., Varesano, N., Pizzutillo, P., Galetta, D., Borelli, M., Barbieri, P., Licen, S., & de Gennaro, G. (2020). Breath analysis: Comparison among methodological approaches for breath sampling. *Molecules (Basel Switzerland)*, 25(24), 5823. <https://doi.org/10.3390/molecules25245823>
- Doran, S. L. F., Romano, A., & Hanna, G. B. (2017). Optimisation of sampling parameters for standardised exhaled breath sampling. *Journal of Breath Research*, 12(1), 016007. <https://doi.org/10.1088/1752-7163/aa8a46>
- Drabińska, N., Flynn, C., Ratcliffe, N., Belluomo, I., Myridakis, A., Gould, O., Fois, M., Smart, A., Devine, T., & Costello, B. D. L. (2021). A literature survey of all volatiles from healthy human breath and bodily fluids: The human volatilome. *Journal of Breath Research*, 15(3). <https://doi.org/10.1088/1752-7163/abf1d0>
- Duizer, C., & de Zoete, M. R. (2023). The role of microbiota-derived metabolites in colorectal cancer. *International Journal of Molecular Sciences*, 24(9), 8024. <https://doi.org/10.3390/ijms24098024>
- Ferrandino, G., De Palo, G., Murgia, A., Birch, O., Tawfike, A., Smith, R., Debram-Beecham, I., Gandelman, O., Kibble, G., Lydon, A. M., Groves, A., Smolinska, A., Allsworth, M., Boyle, B., van der Schee, M. P., Allison, M., Fitzgerald, R. C., Hoare, M., & Snowden, V. K. (2023a). Breath biopsy[®] to identify exhaled volatile Organic compounds biomarkers for liver cirrhosis detection. *Journal of Clinical and Translational Hepatology*, 11(3), 638–648. <https://doi.org/10.14218/JCTH.2022.00309>
- Ferrandino, G., Ricciardi, F., Murgia, A., Banda, I., Manhota, M., Ahmed, Y., Sweeney, K., Nicholson-Scott, L., McConville, L., Gandelman, O., Allsworth, M., Boyle, B., Smolinska, A., Frings, G., Contreras, C. A., Asenjo-Lobos, J., Barrientos, C., Clavo, V., Novoa, N., & Méndez, A., L (2023b). Exogenous volatile organic compound (EVOC[®]) breath testing maximizes classification performance for subjects with cirrhosis and reveals signs of portal hypertension. *Biomedicine*, 11(11). <https://doi.org/10.3390/biomedicine11112957>. Article 11.
- Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C., & Sansone, S. A. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3), 175–178. <https://doi.org/10.1007/s11306-007-0070-6>
- Grabowska-Polanowska, B., Skowron, M., Miarka, P., Pietrzycka, A., & Śliwka, I. (2017). The application of chromatographic breath analysis in the search of volatile biomarkers of chronic kidney disease and coexisting type 2 diabetes mellitus. *Journal of Chromatography B*, 1060, 103–110. <https://doi.org/10.1016/j.jchromb.2017.05.030>
- Haworth, J. J., Pitcher, C. K., Ferrandino, G., Hobson, A. R., Pappan, K. L., & Lawson, J. L. D. (2022). Breathing new life into clinical testing and diagnostics: Perspectives on volatile biomarkers from breath. *Critical Reviews in Clinical Laboratory Sciences*, 59(5), 353–372. <https://doi.org/10.1080/10408363.2022.2038075>
- Herbig, J., & Beauchamp, J. (2014). Towards standardization in the analysis of breath gas volatiles. *Journal of Breath Research*, 8(3), 037101. <https://doi.org/10.1088/1752-7155/8/3/037101>
- Hewitt, M. J., Belluomo, I., Zuffa, S., Boshier, P. R., & Myridakis, A. (2022). Variation of volatile organic compound levels within ambient room air and its impact upon the standardisation of breath sampling. *Scientific Reports*, 12(1), 15887. <https://doi.org/10.1038/s41598-022-20365-7>
- Issitt, T., Wiggins, L., Veysey, M., Sweeney, S. T., Brackenbury, W. J., & Redeker, K. (2022). Volatile compounds in human breath: Critical review and meta-analysis. *Journal of Breath Research*, 16(2), 024001. <https://doi.org/10.1088/1752-7163/ac5230>
- Jia, Z., Patra, A., Kutty, V. K., & Venkatesan, T. (2019). Critical review of volatile organic compound analysis in breath and in vitro cell

- culture for detection of lung cancer. *Metabolites*, 9(3), 52. <https://doi.org/10.3390/metabo9030052>
- Kuo, T. C., Tan, C. E., Wang, S. Y., Lin, O. A., Su, B. H., Hsu, M. T., Lin, J., Cheng, Y. Y., Chen, C. S., Yang, Y. C., Chen, K. H., Lin, S. W., Ho, C. C., Kuo, C. H., & Tseng, Y. J. (2020). Human breathomics database. *Database: The Journal of Biological Databases and Curation*, 2020, baz139. <https://doi.org/10.1093/database/baz139>
- Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, 19(1), 29–41. <https://doi.org/10.1111/1462-2920.13589>
- Majumdar, A., Siva Venkatesh, I. P., & Basu, A. (2023). Short-chain fatty acids in the microbiota-gut-brain axis: Role in neurodegenerative disorders and viral infections. *ACS Chemical Neuroscience*, 14(6), 1045–1062. <https://doi.org/10.1021/acscemneuro.2c00803>
- Maurer, F., Wolf, A., Fink, T., Rittershofer, B., Heim, N., Volk, T., Baumbach, J. I., & Kreuer, S. (2014). Wash-out of ambient air contaminations for breath measurements. *Journal of Breath Research*, 8(2), 027107. <https://doi.org/10.1088/1752-7155/8/2/027107>
- Miller, T. L., & Wolin, M. J. (1996). Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Applied and Environmental Microbiology*, 62(5), 1589–1592. <https://doi.org/10.1128/aem.62.5.1589-1592.1996>
- Mochalski, P., King, J., Mayhew, C. A., & Unterkofler, K. (2023). A review on isoprene in human breath. *Journal of Breath Research*, 17(3). <https://doi.org/10.1088/1752-7163/acc964>
- Mochalski, P., King, J., Unterkofler, K., & Mayhew, C. A. (2024). Unravelling the origin of isoprene in the human body—a forty year Odyssey. *Journal of Breath Research*, 18(3). <https://doi.org/10.1088/1752-7163/ad4388>
- Ney, L. M., Wipplinger, M., Grossmann, M., Engert, N., Wegner, V. D., & Mosig, A. S. (n.d.). Short chain fatty acids: Key regulators of the local and systemic immune response in inflammatory diseases and infections. *Open Biology*, 13(3), 230014. <https://doi.org/10.1098/rsob.230014>
- Parada Venegas, D., De la Fuente, M. K., Landskron, G., González, M. J., Quera, R., Dijkstra, G., Harmsen, H. J. M., Faber, K. N., & Hermoso, M. A. (2019). Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in Immunology*, 277.
- Pham, Y. L., Holz, O., & Beauchamp, J. (2023). Emissions and uptake of volatiles by sampling components in breath analysis. *Journal of Breath Research*, 17(3), 037102. <https://doi.org/10.1088/1752-7163/acce34>
- Phillips, M. (1997). Method for the collection and assay of volatile organic compounds in breath. *Analytical Biochemistry*, 247(2), 272–278. <https://doi.org/10.1006/abio.1997.2069>
- Pitcher, C. K., Farmer, A. D., Haworth, J. J., Treadway, S., & Hobson, A. R. (2022). Performance and interpretation of hydrogen and methane breath testing impact of North American consensus guidelines. *Digestive Diseases and Sciences*, 67(12), 5571–5579. <https://doi.org/10.1007/s10620-022-07487-8>
- Preter, V. D., Machiels, K., Joossens, M., Arijis, I., Matthys, C., Vermeire, S., Rutgeerts, P., & Verbeke, K. (2015). Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut*, 64(3), 447–458. <https://doi.org/10.1136/gutjnl-2013-306423>
- Pugliese, G., Trefz, P., Weippert, M., Pollex, J., Bruhn, S., Schubert, J. K., Miekisch, W., & Sukul, P. (2022). Real-time metabolic monitoring under exhaustive exercise and evaluation of ventilatory threshold by breathomics: Independent validation of evidence and advances. *Frontiers in Physiology*, 13, 946401. <https://doi.org/10.3389/fphys.2022.946401>
- Read, N. W., Al-Janabi, M. N., Bates, T. E., Holgate, A. M., Cann, P. A., Kinsman, R. I., McFarlane, A., & Brown, C. (1985). Interpretation of the breath hydrogen profile obtained after ingesting a solid meal containing unabsorbable carbohydrate. *Gut*, 26(8), 834–842.
- Sachdev, A. H., & Pimentel, M. (2013). Gastrointestinal bacterial overgrowth: Pathogenesis and clinical significance. *Therapeutic Advances in Chronic Disease*, 4(5), 223–231. <https://doi.org/10.1177/2040622313496126>
- Sasiene, Z. J., LeBrun, E. S., Schaller, E., Mach, P. M., Taylor, R., Candelaria, L., Glaros, T. G., Baca, J., & McBride, E. M. (2024). Real-time breath analysis towards a healthy human breath profile. *Journal of Breath Research*, 18(2). <https://doi.org/10.1088/1752-7163/ad1cfl>
- Schubert, J. K., Miekisch, W., Birken, T., Geiger, K., & Nöldge-Schomburg, G. F. E. (2005). Impact of inspired substance concentrations on the results of breath analysis in mechanically ventilated patients. *Biomarkers: Biochemical Indicators of Exposure Response and Susceptibility to Chemicals*, 10(2–3), 138–152. <https://doi.org/10.1080/13547500500050259>
- Spaněl, P., Dryahina, K., & Smith, D. (2013). A quantitative study of the influence of inhaled compounds on their concentrations in exhaled breath. *Journal of Breath Research*, 7(1), 017106. <https://doi.org/10.1088/1752-7155/7/1/017106>
- Sukul, P., Richter, A., Junghanss, C., Schubert, J. K., & Miekisch, W. (2023). Origin of breath isoprene in humans is revealed via multi-omic investigations. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-05384-y>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- van Vorstenbosch, R., Cheng, H. R., Jonkers, D., Penders, J., Schoon, E., Masclee, A., van Schooten, F. J., Smolinska, A., & Mujagic, Z. (2023). Systematic review: Contribution of the gut microbiome to the volatile metabolic fingerprint of colorectal neoplasia. *Metabolites*, 13(1). <https://doi.org/10.3390/metabo13010055>
- Wang, J., Zhu, N., Su, X., Gao, Y., & Yang, R. (2023). Gut-microbiota-derived metabolites maintain gut and systemic immune homeostasis. *Cells*, 12(5), 793. <https://doi.org/10.3390/cells12050793>
- Westhoff, M., Friedrich, M., & Baumbach, J. I. (2022). Simultaneous measurement of inhaled air and exhaled breath by double multicapillary column ion-mobility spectrometry, a new method for breath analysis: Results of a feasibility study. *ERJ Open Research*, 8(1). <https://doi.org/10.1183/23120541.00493-2021>
- Westphal, K., Dudzik, D., Waszczuk-Jankowska, M., Graff, B., Narkiewicz, K., & Markuszewski, M. J. (2022). Common strategies and factors affecting off-line breath sampling and volatile organic compounds analysis using thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). *Metabolites*, 13(1), 8. <https://doi.org/10.3390/metabo13010008>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.