

This is a repository copy of *Volatile compounds in human breath : critical review and meta-analysis*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/183477/>

Version: Accepted Version

Article:

Issitt, Theo, Wiggins, Laura, Veysey, Martin et al. (3 more authors) (Accepted: 2022)
Volatile compounds in human breath : critical review and meta-analysis. Journal of Breath Research. ISSN 1752-7155 (In Press)

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ACCEPTED MANUSCRIPT • OPEN ACCESS

Volatile compounds in human breath: critical review and meta-analysis

To cite this article before publication: Theo Issitt *et al* 2022 *J. Breath Res.* in press <https://doi.org/10.1088/1752-7163/ac5230>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

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

1 2 3 4 5 **Volatile compounds in human breath: critical** 6 **review and meta-analysis** 7 8 9

10
11 **Theo Issitt¹, Laura Wiggins¹, Martin Veysey^{2,3}, Sean T. Sweeney^{1,2}, William J.**
12 **Brackenbury^{1,2} and Kelly Redeker¹**
13
14

15
16 ¹ Department of Biology and ²York Biomedical Research Institute, University of York,
17 York, YO10 5DD, UK, ³School of Medicine & Public Health, University of Newcastle,
18 Callaghan, NSW, Australia.
19

20 Corresponding author:

21 Kelly Redeker, kelly.redeker@york.ac.uk, Tel: +44(01904) 328560, YO10 5DD
22
23
24

25 **Conflict of interest:** The authors declare no conflict of interest
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 **Significance “take-home message”**

54 Human breath holds the potential for accurate disease diagnosis but human breath is
55 a complex mixture, reducing diagnostic efficacy. This work reviews and analyses
56 published outcomes, revealing promising new approaches for future studies.
57
58
59
60

Abstract

Volatile compounds contained in human breath reflect the inner workings of the body. A large number of studies have been published that link individual components of breath to disease, but diagnostic applications remain limited, in part due to inconsistent and conflicting identification of breath biomarkers. New approaches are therefore required to identify effective biomarker targets. Here, volatile organic compounds have been identified in the literature from four metabolically and physiologically distinct diseases and grouped into chemical functional groups (e.g. – methylated hydrocarbons or aldehydes; based on known metabolic and enzymatic pathways) to support biomarker discovery and provide new insight on existing data. Using this functional grouping approach, principal component analysis doubled explanatory capacity from 19.1% to 38% relative to single individual compound approaches. Random forest and linear discriminant analysis reveal 93% classification accuracy for cancer. This review and meta-analysis provides insight for future research design by identifying volatile functional groups associated with disease. By incorporating our understanding of the complexities of the human body, along with accounting for variability in methodological and analytical approaches, this work demonstrates that a suite of targeted, functional volatile biomarkers, rather than individual biomarker compounds, will improve accuracy and success in diagnostic research and application.

Introduction

Human breath analysis offers a diagnostic tool that is non-invasive, rich in information, and low cost. Identification of the presence and abundance of gaseous biomarkers offers the potential for sensitive and accurate clinical diagnosis and long-term monitoring [1–3]. Our ability to ‘translate’ these signals into usable diagnostics currently lags behind the body of published research on captured breath compounds. Despite challenges faced in human breath research, quantification of individual compounds is already used to identify (mal)function of bodily processes in limited contexts. A major challenge of this developing field is aligning volatile compounds captured from breath with underlying (patho)physiologies. In particular, human breath-based clinical trials data is currently insufficiently integrated with our understanding of functional and mechanistic physiology. The focus here is on human breath, but gaseous biomarkers can be detected from skin, urine, blood, saliva and faeces [4,5].

The majority of breath-linked diagnostic research has targeted respiratory diseases. Attempts to identify volatile organic compound (VOC) biomarkers of lung cancer, both *in vitro* and *in vivo* [6–8] are represented by a large body of work. Non-cancerous pulmonary diseases such as asthma [9,10], chronic obstructive pulmonary disease (COPD) [11,12], cystic fibrosis [13,14] and tuberculosis [15,16] are also targets of research, but to a lesser extent.

In addition to pulmonary disease, VOC biomarkers from other cancers [17], cardiac disease [18], liver [19], gastrointestinal [20], and neurological conditions [21,22] have been studied and reported. The breadth of these studies offers an opportunity to compare how variable cellular states and pathophysiology correlate and/or differ in

VOC profile. For example, diabetically linked VOCs [23] give insight into metabolic functions that may have implications for other disease-correlating phenotypes.

The diagnosis of infection is a promising field for breath research, in part because microorganisms often generate distinct VOCs, which can be discerned within human breath profiles [24,25]. For example; tuberculosis [16,26] and *Pseudomonas aeruginosa* [27,28], both infections of the lung, are metabolically distinct. Viral infections, separate from microbial infections, may also be detectable due to viral VOC production, or modification of the human host metabolism. Canine detection by scent of COVID-19 has been demonstrated [29,30] and several studies have reported on the potential for COVID-19 breath-based diagnosis [31–37], with varying accuracies [38].

Differences between disease states can increase the power of diagnostic tests. Bayes' Theorem links probability of disease to prevalence within a population as well as the presence or absence of clinical markers [39]. This paper aims to highlight the need for diagnostic research frameworks that include VOC biomarkers which act as comparative controls to increase diagnostic precision and accuracy.

VOCs - The complex pathway from cell to breath

Human breath contains hundreds of volatile organic compounds. Metabolic processes within the human body both consume and generate VOCs, also referred to as the 'volatilome'- which is defined as the volatile fraction of the metabolome [4]. As a fraction of the metabolome, VOCs are recognized to directly reflect gene transcription and protein expression. Illness, which is often linked to altered metabolisms and local environmental changes, is therefore expected to alter 'volatilome' profiles. The available human 'volatilome' consists of gaseous, low concentration ($<1 \times 10^{-4}$ %), low molecular weight (<350 amu molecular weight), and high-vapour pressure compounds extant within the gas phase at human temperatures and ambient pressures.

The primary target of most breath research are endogenous (internally generated) VOCs however, human breath consists of a mixture of both endogenous and exogenous VOCs. Exogenous VOCs arise from sources external to the body which include local air volatiles (e.g. - car exhaust) as well as metabolic by-products from diet and/or medications. Exogenous compounds that persist continually in the environment (i.e. the clinic, or urban streets) must be characterised, quantified and separated in order to clarify which endogenous compounds are produced or metabolised by the patient. Quantifying metabolism of exogenous VOCs can be a powerful diagnostic tool in its own right- for example in organ function, where metabolism of limonene (e.g. - produced by air fresheners and various plants) can be used to assess liver function [40]. This technology can be applied to preoperative and postoperative assessment of liver function, and drug-induced liver damage [41]. Alternatively, through utilisation of easily detectable, stable isotopically labelled molecules, such as 13 -Carbon labelled hydrocarbons, specific bodily processes can be monitored and assessed through breath [42] including measurement of gastric emptying through labelled CO_2 levels [43] or labelled Urea in the breath, indicative of *H. pylori* [44,45]. Similarly, levels of hydrogen in the breath can accurately assess malabsorption in the gastric tract through bacterial processing of administered fructose [46,47].

1
2
3 Further to the use of exogenous compounds as molecular probes, some of the most
4 impactful and fundamental breath research has focused upon the effect of exogenous
5 VOCs on human health, increasing understanding of volatile dynamics [48]. For
6 example, the effects of cigarette smoke and carcinogenic VOCs [49] or exposure to
7 VOCs in firefighters [50]. Some of the VOCs outlined here as biomarkers of disease
8 have been identified as toxic to human health, such as benzene, 1,3-butadiene,
9 styrene and isoprene, the most abundant VOC in human breath [51]. Therefore
10 concentration and context is an important factor when investigating volatile dynamics.
11
12

13 Microbial emissions also produce quasi-exogenous volatiles that may be revealing of
14 pathological conditions or confound diagnosis. Cells of non-human origin outnumber
15 the body's cells by far [52] and their metabolisms form a considerable fraction of the
16 VOCs released and metabolised by the human body [4,53]. For example, VOCs like
17 acetone can be produced by anaerobic and aerobic bacteria [54,55] and residual
18 levels of ethanol and methanol can be either exogenous or microbial in origin [56].
19 Usefulness of volatile biomarkers is therefore defined by pathophysiology and
20 comorbidities since altered microbiomes may be a significant and defining source of
21 VOCs. This is especially likely in disorders of the bowel [20,57].
22
23
24

25 Breath analysis therefore requires systemic approaches for successful diagnostic
26 application, accounting for both patient variability and environmental effects. Increased
27 understanding of volatile metabolic processing in the body will aid in contextualising
28 the qualitative and quantitative effects of stress, age, time of day, gender, activity,
29 disease status, and/or transport of VOCs to the site of detection, all of which affect
30 VOCs in breath [2,8].
31
32

33 VOCs produced or metabolised by cellular processes, which are not subject to direct
34 diffusion to exhaled air, must travel around the body through the bloodstream. They
35 will i) pass tissues with varying constitutions and affinities to that volatile, ii) be
36 metabolised through enzymatically independent and dependent pathways (the
37 majority of these enzymes are expressed in the liver), and iii) diffuse from the
38 bloodstream into the lung air space across the alveolar wall (Figure 1). In the lung,
39 VOCs released from the blood mix with all local metabolites and metabolising agents
40 prior to exhalation. In the mouth volatiles from the lung, mouth, nose, upper
41 gastrointestinal tract and stomach mix prior to sampling.
42
43

44 The primary physicochemical properties governing VOC movement within the body
45 are blood:air and lipid:air partitioning coefficients, representing how likely a volatile is
46 to solubilize in aqueous solutions (e.g. - blood) or dissolve into fat [58]. These basic
47 thermodynamic properties are governed in human tissues by a molecule's size and
48 polarity. For most cells in the body, volatiles initially move between the blood and the
49 cell rather than directly diffusing into alveolar airspace (Figure 1). Once a volatile
50 compound enters the bloodstream it must pass through the organs and tissues of the
51 body for which it may have variable affinities [58,59]. The relative distribution of tissue
52 types is a major source of variability between individuals. Lipophilic volatiles can
53 accumulate in fat tissues, while compounds with low affinity for fat drain more
54 efficiently into the blood, highlighting body mass index effects on VOCs in breath [2].
55
56
57

58 Further metabolism of released VOCs within the body can substantively modify the
59 available volatilome. For instance, altered metabolisms (e.g.- the state of ketosis or
60 fasting) have been shown to alter breath VOC profile [60]. In diseases such as

1
2
3 diabetes, changes in acetone can be indicative of diabetic ketoacidosis [61]. This
4 global change, affecting all cells, contrasts with VOC sinks and/or sources that are site
5 specific such as tumours, whereby the local microenvironment may present alterations
6 in pH [62], hypoxia [63] and cellular ion concentrations [64].
7

8
9 It is important to note that individual VOC biomarkers linked to cellular state may not
10 be able to differentiate between causative agents or symptoms. For example; cellular
11 iron overload [65], senescence [66] or cell death [67] may all produce mitochondrial
12 dysfunction and oxidative stress, producing similar VOC biomarkers. Therefore, VOC
13 research should aim to identify differences in volatile metabolic outcomes between
14 these states that may translate to the breath.
15

16 Challenges in VOC biomarker comparison and collection

17
18 Volatile-focused biomarker research is confounded by varying behaviours and
19 metabolisms between individuals [2]. Volatile biomarkers may well be indicative of an
20 isolated cell, *in vitro*, but within a body, may be subject to further metabolism (Figure
21 1). Some VOCs may therefore be a direct readout of enzymatic activity while others
22 reflect multiple enzymatic processes. For example, limonene, which is not produced
23 by human metabolism, can be measured in breath to monitor liver function as a read
24 out of cytochrome P450 (CYP450) activity [19,40]. Whereas, peroxidation of lipids,
25 hypothesised to be a source of aldehydes and hydrocarbons in the breath [4,23,68,69],
26 can be mediated by enzymes, such as lipoxygenase, cyclooxygenases or cytochrome
27 P450 [70] or non-enzymatic peroxidation through oxygen-radical oxidative routes [71].
28
29

30
31 The direct processing of functional VOC groups, such as aldehydes, makes several
32 enzymes both sources of VOC biomarkers and potential confounding elements in the
33 processing of VOCs produced from other cellular mechanisms. Some of these
34 enzymes should be considered as confounding factors affecting translation of
35 research, as they may break down primary targets. Of this wide array of enzymes,
36 Alcohol Dehydrogenases (ADHs), Aldehyde Dehydrogenases (ALDHs), Aldehyde
37 Oxidases (AOX), Aldo-keto reductases (AKRs) and Short-chain
38 dehydrogenases/reductases (SDRs) are examples of classes that directly influence
39 commonly detected and targeted VOCs.
40
41

42
43 Once endogenous metabolism and associated differences in form/behaviour have
44 been addressed there are still methodological approaches which can bias reported
45 outcomes. The pro's and con's of various techniques have been reviewed previously
46 in the context of human breath [6,72]. Briefly, variability in reported information can
47 occur through analytical approach (i.e.- the instrument through which the data is
48 quantified) [72] or in sampling approach. Sampling modifies reported results through
49 changes in temperature, humidity, phase of breath (alveolar vs whole breath) or
50 expiratory flow rate [6].
51
52

53
54 Analytical and collection methodologies vary across published studies (Table S1). In
55 studies where breath samples are taken and concentrated prior to analysis, most
56 studies collect breath into specialist polymer bags or use chemical traps. Collection
57 methods should be considered when collecting and interpreting as they can affect the
58 VOCs which are collected. For example, Tedlar® bags can affect breath VOCs through
59 compound degradation and interaction with the bag product [73–76]. On the whole,
60

1
2
3 most studies use some form of thermal desorption tube (TD) containing a specialized
4 sorbent or solid phase microextraction (SPME) fibre (Tables S1/S2). Each available
5 suite of methods results in compound bias. While researchers attempt to counter such
6 biases, methodological variability inevitably generates inconsistency in published VOC
7 outcomes.
8
9

10 Variation in reported human breath outcomes, and associated biomarkers, therefore
11 results from;
12

- 13 1. Variability inherent in, and between, sampling methodologies;
- 14 2. Inherent human variability;
- 15 3. Complex interactions between compounds in breath; and
- 16 4. Confounding signals from comorbidities.
17
18

19 Like any diagnostic tool, precise and accurate interpretation of results depends on our
20 ability to statistically link detectable changes to outcomes. Due to the complexities that
21 arise from varying individual metabolisms and variability derived from methodological
22 approaches, volatile biomarkers have been inconsistently reported, in terms of both
23 presence/absence and quantity, for a range of diseases. For example, propanol,
24 isoprene, acetone, pentane, hexanal, benzene, ethylbenzene and toluene have
25 individually been reported to be lung cancer identifiers in 6 or more studies [6].
26 However, increases in isoprene (as one example) from lung cancer patient breath
27 compared to control groups [77–80] conflicts with reports where isoprene decreased
28 in lung cancer patients [81,82]. To date, published diagnostic compounds from human
29 breath appear to demonstrate little continuity with *in vitro* models [6].
30
31
32

33 Having outlined the challenges faced by researchers in identification of volatile
34 biomarkers in breath, in this paper, we perform a comparative analysis that will allow
35 researchers to identify and target biomarkers linked to pathophysiology and to
36 consider their work in the context of a range of human diseases. Through considering
37 how disease location, VOC interaction, and systemic variability affects end-point
38 breath profiles, research efforts can be more clearly focussed and optimised.
39
40
41

42 Methodological approach and rationale

43 Breath research is varied and multiple tools and approaches for research exist. Some
44 of the most active areas of breath research include; lung cancer, breast cancer,
45 cancers of the mouth, throat and upper gastrointestinal tract, diabetes, liver disease
46 and inflammatory bowel disease (IBD). Our collation of this data is based on available
47 studies, with lung cancer studies outnumbering all other breath research.
48
49

50 Volatile biomarker comparison and data collection

51 To demonstrate the challenges researchers face in deriving VOC biomarkers from
52 breath research we collected data from four metabolically and physiologically distinct
53 diseases for which there exist a number of available studies (Figure 2, Table S1).
54 Several systematic and comprehensive reviews; for lung, breast and other cancers
55 [6,83], irritable bowel studies [84,85], diabetes [23] and liver disease [19] were cross
56 referenced to widen scope and inclusion of studies (see PRISMA flow chart,
57 supplementary figure 1). Detailed exclusion criteria, workflow and transparent
58
59
60

1
2
3 methodology can be seen in supplementary and in the PROSPERO database [86].
4 Systematic searches of title and abstract were performed for each disease using
5 boolean operators AND and OR using the Embase and Medline databases through
6 the OVID platform. Detailed information on systematic search terms and results are
7 provided in supplementary methods along with PRISMA flow chart [87]. This research
8 can also be found on the PROSPERO data base [86] where clear inclusion criteria,
9 methodology and data extraction are given. Risk of bias and data analysis can also be
10 found in supplementary materials.
11
12

13 It is important to note that a range of important studies into the breath of patients with
14 pulmonary disease such as asthma [9,10], COPD [11,12,88], cystic fibrosis [13,14]
15 and tuberculosis [15] as well as many other diseases, infectious or otherwise have
16 been conducted. However, for these diseases, there do not yet exist sufficient studies
17 fitting the selection criteria for inclusion here. This is highlighted by a systematic review
18 into breath analysis and COPD [88], which identifies 12 papers, many of which use
19 smokers as a control group, and highlights the lack of clinical breath biomarkers [86].
20 Neurodegenerative disease also shows promise as a breath diagnostic application,
21 but it is still a developing field and more biomarker research into breath needs to be
22 conducted. Infectious diseases also suffer from this same problem with the added
23 element of many different types being investigated, making them incompatible for this
24 meta-analysis. Asthma, COPD and parkinsons disease have also been searched
25 using our methodology, and the results discussed in the supplementary.
26
27
28

29 Reviewer T.I. screened outcomes from electronic searches, their inclusion was based
30 on criteria outlined in supplementary methods. This was double checked by reviewer
31 K.R.
32
33

34 Separation of studies based on methodology

35 Pilot studies are often employed when investigating VOCs in human breath of
36 diseases that have not been investigated before [21,89–94]. These studies examine
37 compounds in breath using a more untarged or scanning approach (identified here as
38 SCAN). Compared to control groups, statistically significant increases in VOCs often
39 form the basis and rationale for investigating identified compounds at more depth in
40 future studies.
41
42

43 Non-targeted (discovery), or scanning, approaches to gas analyses are useful for
44 identifying where signals are substantially altered when compared to control groups.
45 However, the large number of compounds in breath (>1000 [4,5,95]) and the (usually)
46 single temporal sampling approach often means that only compounds that exceed
47 substantial signal-to-noise ratios, constrained by sampling and analysis methods, and
48 that overcome complexities associated with individual and population variability are
49 reported. Informative compounds that exist in smaller quantities or compounds that
50 are absorbed and metabolised by the body are often missed from these types of
51 scanning studies and therefore subsequent targeted selective-ion mode (SIM)
52 analyses may be searching in the wrong place.
53
54
55

56 Of the studies utilised in this meta-analysis, only those studies where compounds were
57 reported as increased, when compared to control group, have been used. Where
58 studies have reported VOC uptake, they have not been included due to the rarity of
59 this approach. This is a significant lapse in the published literature as volatile uptake
60

1
2
3 may form a very important avenue for biomarker discovery. When combined with
4 longitudinal studies in diagnostic applications, this approach may help to overcome
5 systemic issues affecting cohort variability.
6

7
8 Targeted, or Selected/Selective Ion Monitoring (SIM), MS analyses can provide a
9 more sensitive, targeted approach to quantifying volatiles in breath. By focusing on
10 individual compounds, researchers achieve substantially greater methodological
11 sensitivity in detection and quantification. Monitoring targeted VOCs can provide in-
12 depth information about complex processes, such as the citric acid cycle [96] but it is
13 important to focus on correct VOCs for accurate diagnosis.
14

15
16 In the first stage of data collection, papers were considered regardless of detection
17 method (Table S1). For further analysis, studies were reduced to SCAN studies as
18 well as those studies which searched for a suite of volatiles that were representative
19 of multiple functional groups (Table S2).
20

21 Categorical variables

22
23 Of the 84 studies retained after selection criteria, 43 focus on lung cancer, five breast
24 cancer, 13 Diabetes, 13 liver disease and eight IBD. Five further studies that focused
25 on cancers of the stomach, mouth, larynx and upper gastrointestinal region, are
26 grouped as upper gastrointestinal (UGI) cancers. It is outside the scope of this
27 research to consider pathological variability within each group, due to limited study
28 numbers, therefore diseases have been grouped. For example; the liver disease group
29 includes studies investigating liver cirrhosis in adults and non-alcoholic fatty acid liver
30 disease (NAFLD) in children. Variability in pathology has been noted in Table S2.
31 Furthermore, diagnosis and separation of pathologies may cause inaccuracies when
32 using breath volatiles, for example, separation of IBD conditions: Crohn's disease and
33 Ulcerative colitis from control groups can be accurate but separating the two
34 pathological profiles is less accurate [97].
35
36
37

38 Studies investigating limited numbers of compounds generate uninformative
39 outcomes when compared with studies investigating different, targeted compounds or
40 studies employing a non-targeted approach (Table S1). However, some variability is
41 likely to be due to the range of instrumental and collection techniques employed. Most
42 studies listed here utilised Gas Chromatography Mass-Spectrometry (GC/MS) as their
43 analytical platform, but other methods include Proton-Transfer Reaction Mass-
44 Spectrometry (PTR-MS), Selected Ion Flow-Tube (SIFT-MS), Ion-Mobility-
45 Spectrometry (IMS) (Table S1). There may also be further subdivisions, for example,
46 standard GC/MS or GCxGC TOF, all of which will have an impact upon the observation
47 of compounds [98]. These methods should be considered when comparing reported
48 VOCs between studies.
49
50
51

52 Meta-Analysis and compound nesting

53
54 For each study, reviewer TI extracted data of VOCs which were identified as
55 increased/enhanced in concentration. Volatiles reported from these studies were
56 compared through Principal Component Analysis (PCA) using a binary function -
57 present (1) or not (0) in a matrix (Table S2) using R-studio and ggplot2. This data was
58 then used to train two classification models, random forest (RF) and linear discriminant
59 analysis (LDA), with predictions and classification accuracy scores obtained through
60

1
2
3 leave-one-out cross validation. All classification was performed in R-studio, using the
4 randomForest package for RF, and the MASS package for LDA. Complete equal
5 weighting of individual compounds and/or equal consideration of all possible individual
6 VOCs (i.e when considering each possible compound) led to uninformative PCA
7 outcomes. Compound nesting (combining similar/related compounds under one
8 heading) was applied to clarify PCA outcomes. As an example of compound nesting;
9 monomethylated alkanes, such as methylated variants of undecane (of which 4
10 isomers exist), have been considered as one VOC biomarker in Figure 2. The nesting
11 categories can be seen in supplementary tables.
12
13
14

15 Results

16
17 By considering every reported biomarker across a wide variety of studies and
18 methodologies, including targeted single biomarker studies, no single or suite of VOC
19 compounds show diagnostic potential for lung cancer (Figure 2). Primary PCA axes
20 explain very little of the observed variance across all studies (only 19.1% of the
21 variation within the data can be explained by PCA axes 1 and 2). IBD, diabetes and
22 liver disease are inseparable (within the 95% confidence intervals (CI) assigned by the
23 PCA) while lung cancer VOCs overlap all groups with several outlying studies (Figure
24 2). The lack of definitive outcomes when comprehensively including all reported data,
25 as represented in the PCA analysis (Figure 2), is expected when considering
26 comorbidities, systemic variability and methodological differences.
27
28
29

30 Alternative grouping of compounds may be more informative of processes underlying
31 production of individual compounds. It may be more appropriate to consider, for
32 example, all five carbon alkanes (e.g. pentane or methylated butanes), regardless of
33 methylation or ethylation, as indicative of a functional process, inclusive of modification
34 events. These aggregated 5 carbon compounds may therefore be more descriptive of
35 specific metabolisms than individual compounds, reducing the impacts of individual
36 variability in compound metabolism or derivatization.
37
38

39 We hypothesised that applying a nesting approach (combining compounds with similar
40 functional grouping) would reveal distinctive trends between pathologies in VOC data.
41 For example, altered levels of alcohols in the breath have been reported often in liver
42 disease, including ethanol, methanol and propanol [99]. Similarly, a number of
43 aldehydes have been reported for several cancer types, with hexanal being the most
44 commonly reported [17]. Assuming that hexanal (in cancer) or ethanol (in liver disease)
45 are the critical, important breath biomarkers and not aldehyde/alcohol metabolisms
46 more generally may reduce the information that can be gleaned from single
47 biomarkers. While individual VOC biomarkers may increase specificity, there is a need
48 to perform further analysis to identify how they relate to functional chemical groups
49 and disease/stress-based metabolisms.
50
51
52

53 Functional grouping of Volatiles

54 Data (barring targeted studies) was grouped into chemical functional groups, as
55 defined in table 1 [100]. PCA analysis using functional groups was able to explain
56 substantially more of the data presented (38.0% from axes 1 and 2, Figure 3) and
57 created a clear separation between lung cancer and all other disease states (Figure
58 3). Primary functional groups which separate lung cancer from other diseases include
59
60

1
2
3 hydrocarbons (notably six carbon compounds and above, irrespective of saturation or
4 branching), aldehydes, furans, cyclic hydrocarbons and aromatics (Figure 3). Benzene
5 derivatives (aromatics) were reported in the breath in every lung cancer study.
6 Isoprene, a commonly reported biomarker for cancer [83], has not been included in
7 this analysis due to high variability of published outcomes [101]. Its inclusion however
8 did not significantly alter PCA outcomes.
9

10
11 Most diabetic studies were defined by the appearance of ketones in the breath, notably
12 and unsurprisingly acetone, a volatile commonly associated with diabetes[23] as well
13 as alcohols, including butanol, methanol and ethanol.
14

15
16 IBD was defined by the presence of hydrocarbons (notably, shorter compounds, eight
17 carbons or less), nitrogen and sulphur compounds. The pathophysiology of IBD, such
18 as Crohn's disease is characterized by periodic inflammation (linked to oxidative stress
19 and subsequent hydrocarbon release [102,103]) and an altered microbiome (linked to
20 alterations in sulphur metabolism and nitrogen compounds [19,99]).
21

22
23 Studies investigating forms of liver disease, including NAFLD and cirrhosis, were
24 strongly defined by the presence of monoterpenes in the breath, notably limonene and
25 pinene. However, this was slightly skewed in the PCA analysis as only 4 out of the 10
26 included studies reported monoterpenes (Table S2). Ketones, nitrogen and sulphur
27 compounds were also seen in patients suffering from liver disease. Interestingly, only
28 4 studies reported alcohols in the breath of liver disease patients (one focusing on
29 NAFLD [99], one comparing between NAFLD and cirrhosis [104], and two investigating
30 cirrhosis only, Table S2) and this was less defining of liver disease as a group than
31 other functional groups, despite impaired alcohol processing being a hallmark of liver
32 disease and therefore purportedly a breath biomarker [19].
33
34

35 Cancer comparisons

36
37 To investigate the possibility that grouped lung cancer breath VOC outcomes (Figure
38 3a) were the result of proximity to pulmonary architecture, facilitating direct diffusion
39 of VOCs to lung airspace rather than systemically processed VOCs (Figure 1), we
40 compared VOCs reported in the breath of breast cancer patients and cancers of the
41 upper GI tract and mouth (UGI) to the groups already presented (Figure 3B). Addition
42 of breast and UGI cancer studies reveals close correlation of all cancer groups (breast,
43 lung and UGI) along similar axes, while retaining significant separation from diabetes,
44 liver disease and IBD outcomes. To clearly demonstrate this separation, these
45 subgroups were grouped into cancer vs other (supplementary figure 2A). The PCA
46 biplot is also provided to show which elements were identified as most discriminatory.
47 (supplementary figure 2B). This suggests that PCA separation from other diseases is
48 not due to relative location within the pulmonary architecture, mouth and oesophagus.
49 Associations of functional VOCs are more consistent between cancer pathologies
50 relative to other disease states, suggesting that these signals are disease correlated.
51
52
53

54 Further Analyses

55
56 To further expand on the use of PCA and to acknowledge the presence of potential
57 'voodoo correlations' in the data [105] we performed random forest and LDA for
58 classification of cancerous vs. non-cancerous (other) diseases (figure 4). This
59 demonstrated that the functionally grouped VOCs can be used in combination to
60

1
2
3 classify cancer with high accuracy. Random forest determined 93% accuracy for
4 cancer and 100% for 'other'. LDA determined 93% accuracy for cancer and 81% for
5 'other'.
6
7

8 Discussion

9
10
11 The difficulties faced by breath researchers exploring released VOCs is highlighted by
12 multiple reviews investigating lung cancer and pulmonary disorders [6,69,83,106–
13 109], diabetes, liver disease and IBD [19,23,84,85,110]. Within these reviews, limited
14 consensus has been reached regarding the efficacy of individual compounds to
15 identify specific diseases and/or disease-based metabolisms. These are reviewed in
16 Table 1. In spite of this ongoing and dedicated research, sufficiently precise and
17 accurate breath biomarkers for diagnostic application have continued to elude
18 researchers for cancer [6], liver disease [19], IBD [85] or diabetes [110]. We have re-
19 affirmed the challenges in biomarker identification through our introductory discussion
20 on sources of breath VOC variability and through the lack of descriptive potential, as
21 shown in Figure 2.
22
23

24
25 Functional group analysis, which we show clarifies existing, previously disparate,
26 studies (Figure 3), are not to be taken as recommendations for singular biomarker
27 approaches even when these single biomarkers exist as components of a larger
28 functional group. Clearly, singular volatile approaches are not effective (Figure 2) and
29 this has been recognised by researchers previously [105]. Functional group analysis
30 does, however, provide a guide for further research, described here as a 'breath print
31 and research framework'. We propose that analysis of multiple volatile biomarkers
32 from a range of functional classes will provide increased discriminatory power.
33
34

35 Functional Groups of Volatile Biomarkers

36
37 We have improved disease separation within our PCA analyses (Figure 3a), affirmed
38 that location of disease does not drive reported outcomes (Figure 3b), and highlighted
39 trends in volatiles discovered in human breath by using a functional grouping
40 approach. Successful application of functional groups to biomarker discovery implies
41 that functional groups are more defining of process or disease (Figure 3), than single
42 volatile markers (Figure 2). A recent review has surveyed volatiles released by
43 humans and highlighted functional groups [5].
44
45

46
47 Further to PCA analyses, both random forest and LDA confusion matrices revealed
48 high accuracy in recognising cancer and 'other' (Figure 4). While it should be
49 recognised that grouping very distinct diseases together in this way is confounding
50 within itself, it demonstrates the power of this approach. The classification results
51 obtained here are suggestive that VOCs could prove a powerful tool for cancer
52 diagnostics, with many providing good discrimination between cancerous and non-
53 cancerous diseases.
54

55
56 A number of metabolic pathways and key characteristics of functional groups
57 associated with VOCs in breath have been reviewed [4,23,69] and some of these
58 pathways, pertinent to diseases investigated here, have been highlighted (Table 1).
59 Understanding remains limited and further research into targeted cellular metabolisms
60

is needed to disentangle common functional outcomes. We present here a brief analysis of exogenous sources and endogenous metabolisms for several functional groups with predictive power in our analysis.

Only 4 disease outcomes are included here, due to reasons discussed in the methodology section. For this reason we have included possible pathophysiological sources of compounds, for cross reference to other pathologies in table 1. For example, aldehydes have been linked to inflammatory linked stress and subsequent lipid peroxidation in a range of diseases, such as COPD [109] and can be seen here in cancer (table 1). Possible endogenous sources may then be cross referenced to develop a suite of diagnostic compounds dependent upon disease.

Class	Example Compounds	Prevalent In	Possible Endogenous Source
Hydrocarbon	Butane, Heptane	All	Lipid peroxidation [69,103,111–114], ethanol Metabolism [115]
Alcohol	Ethanol, Propanol	Cancer, diabetes, liver Disease	Alcohol Metabolism, Ketone Metabolism [106,116,117], Hydrocarbon Metabolism [114,118]
Ketone	Acetone, Butanone	Cancer, diabetes, liver Disease	Amino acid metabolism to acetone [17,119–121], isopropanol to acetone [17,122], fatty acid metabolism and oxidation [17,69,123,124]
Aldehyde	Hexanal, Acetaldehyde	Cancer	Lipid Peroxidation [71,103,125,126], alcohol metabolism [127], enzymatic function [128] [17,122]
Carboxylic Acids	Propanoic acid	Cancer (breast)	Aldehyde oxidation [129], Lipid peroxidation [103,130,131], Microbial [132]
Ester/Ether	Butyl Acetate, Dimethyl-Ether	Cancer	Enzymatic action i.e. esterases [69,133]
Isoprenoids	Limonene, Pinene	Liver disease	CYP450 activity [19,40]
Nitrogen	Trimethylamine, Ammonia	IBD, liver Disease	Amino acid metabolism [106], Microbial [134,135]
Furan	Furan	Cancer	Unclear, microbial action [136]
Sulphur	Dimethyl Sulphide, Hydrogen Sulphide	Cancer (lung), IBD, liver Disease	Urea cycle [137], Microbial [19,99]
Aromatic	Benzene, Xylene	Cancer	Released from fatty tissue [138], CYP450 [139], unknown endogenous creation
Cyclic Hydrocarbons/Ketones	Cyclopentane, Cyclohexanone	Cancer	Unclear

Table 1. **Functional groups of volatile compounds seen in breath research studies and possible endogenous sources of variance.** The data presented here links studies (presented in figure 3) to prevalent functional groups of volatiles to cancer types, irritable bowel disease (IBD), diabetes and liver disease.

Aromatics, furans and cyclic hydrocarbons

Cyclic compounds, such as aromatics, cyclic hydrocarbons, and furans, act as important compounds in differentiating between lung cancer and other disease states in our analysis (Figure 3a). Cyclic compounds have, however, generally been regarded as contaminants in breath research [6,69] and, because of this, their diagnostic power has often been dismissed. Due to common exposure as exogenous compounds, the use of these aromatic compounds as diagnostic tools should be taken with caution and use as a single compound diagnostic would not be recommended. They retain diagnostic power however, in part as a negative marker, within our approach.

Benzene (and derivatives) and furans are present in cigarette smoke and higher in the breath of smokers [140], a particular consideration for lung cancer breath profiles. Studies which addressed this by contrasting VOC screens from smokers/non-smokers suffering from lung cancer have found that benzene derivatives and furans were still present [141]. Furthermore, studies have shown that cultured, in vivo cancer cells release a range of benzene derivatives [142–148]. Human fibroblasts [149] and human mammary epithelial cells [143] also produce aromatics when grown in culture. This highlights how false positives from exogenous sources can confound separation of functionally useful markers from contamination.

Furans have been associated with smoking and these compounds are not associated with endogenous origin [150]. Appearance in heated food suggests an association with diet [151]. Furan appears in the breath of healthy, non-smoking individuals in addition to smoking and non-smoking cancer patients and individuals [141,152]. Furans, have been reported in lung cancer, and one study into laryngeal cancer [153] (Table S2). As this compound was not seen in breast cancer or other diseases investigated it suggests that there might be either a) a pulmonary diffusion aspect to detection, b) a smoking component or c) both.

Cyclic hydrocarbons, such as cyclopentane and cyclohexane, have not been investigated with respect to metabolic cellular function but their appearance in the headspace of cell lines such as mesothelioma [154] and their effective use diagnosing cancer patients from breath for colorectal [155], lung [92,156] and breast cancer [157] suggests that they retain diagnostic capacity irrespective of exogenous contaminant sources. Cyclohexanone and other cyclic hydrocarbons are by-products of plastic and fuel combustion [158], and are unlikely to be contaminants in cellular headspace analysis. Cyclohexane has been shown to be descriptive of malignant pleural mesothelioma when contrasted with subjects with similar professional asbestos exposure [159]. However, oxygenation of cyclohexane produces cyclohexanone, thought to be a result of fatty acid oxidation and weight loss [160].

Throughout the data presented here, furans, cyclic hydrocarbons, aromatic compounds and benzene derivatives have been consistent markers of cancer, irrespective of lung cancer, breast cancer or cancers of the mouth and upper GI tract (Figure 3 and Table S2) [161]. While these compounds all have exogenous sources, this work highlights their diagnostic potential. While in many instances, they may be confounded with smoking related diseases, their absence from IBD, liver disease and diabetes studies, may allow diagnosticians to remove these diseases from

1
2
3 consideration, providing a powerful combination of VOC biomarkers and a starting
4 point for comprehensive 'breath print' analyses.
5

6 7 Developing a 'breath-print' and research framework

8
9 The identification of a single volatile biomarker for diagnosis of complex pathologies,
10 appears unlikely considering the unsuccessful outcomes of more than three decades
11 of research on diseases such as lung cancer. It seems more likely that multiple
12 biomarkers will provide maximum diagnostic accuracy and this has been recognised
13 by breath researchers [105,162–164]. For example, acetone has been a target for
14 diabetic breath research since the 1960s [165], linked to ketoacidosis [110] and
15 characteristic of the sweet smell on the breath [166] and found in greater
16 concentrations in the breath of diabetics. However, as a single marker it does not
17 optimise diagnostic potential, due to concentration variability linked to insulin
18 resistance, lipolytic activity, exercise, fasting status and gender [23]. Other VOC
19 markers can therefore be utilised in tandem to build up a 'breath-print', increasing
20 diagnostic power and overcoming systematic variability and comorbidities.
21
22

23
24 In addition to multiple volatile biomarkers increasing diagnostic accuracy, a 'breath-
25 print' may potentially include wider breath dynamics and pulmonary function such as
26 flow rate, pressure, and gas transfer. Combined, this can build an accurate picture of
27 lung function [167] and these tools are used frequently in the clinic for assessing
28 patients with COPD, asthma or any restriction to breathing [168,169]. Lung function
29 impacts testing and collection of volatiles, creating variability between individuals and
30 so consideration of this will increase the power of diagnosis by VOCs.
31

32
33 Recommendations for volatile compounds as disease diagnostic markers have not yet
34 been made for many disease states. This is, in part, due to the variation in approaches
35 (Table S1) and systemic complications (Figure 1). In this research we have arranged
36 reported markers from non-exclusionary studies into functional groups and
37 substantially improved disease separation, generating greater correlation across
38 primary PCA axes (Figure 3). We have also shown that cancer studies generate similar
39 outcomes, irrespective of location, lending credence to the idea that our reported
40 outcomes are independent of bodily location and, therefore, due to common metabolic
41 action. This work agrees with systematic and prospective reviews which have
42 identified correlations between disease compounds such as aldehydes for cancer
43 diagnosis [6,17,83]. We, therefore, recommend that research targets consist of a suite
44 of markers that encompass a range of functional groups.
45
46

47
48 Application of functional group analysis is limited as it can remove specificity. For
49 example, butanone and acetone are both ketones but butanone is highly present in
50 the lung cancer group but not in the diabetic group (Table S2). Therefore, when
51 selecting compounds for investigation, a selection of compounds from several
52 functional groups (i.e. 3 ketones, 3 aldehydes, 3 hydrocarbons, 3 sulphur compounds
53 etc) may optimise descriptive and diagnostic potential.
54

55
56 Accordingly, a suite of VOCs (a 'breath-print') can be utilised to account for variability
57 within individuals. However, an understanding of functional groups and how they relate
58 to metabolic processes will allow for more effective identification of volatile compounds
59 to serve as biomarkers within the 'breath-print'. Based on group separation in figure
60

3b, IBD, liver and diabetes are separable from lung, breast and OG cancer, but each of these sub-groups (cancers versus gut/liver diseases) have a number of overlapping compounds. A suite of VOCs for targeted cancer diagnosis would include both positive and negative markers. The following outlines a framework for developing cancer diagnostic targets for breath where a study might focus on 16 to 20 VOCs.

Positive markers of cancer would include;

- I. Aldehydes, such as; pentanal, hexanal and heptanal;
- II. Multiple hydrocarbons above 6 carbons such as heptane, octane and decane (there appears to be no preference for branched chained hydrocarbons in the data)
- III. Aromatic and cyclic compounds, such as ethyl benzene, furan, cyclopentane and cyclohexanone.

The ketone, butanone, was also highly reported for cancer studies. Presence of each biomarker individually is not confirmation of diagnosis but acts to increase diagnostic accuracy.

Negative markers might include;

- I. Monoterpenes, either limonene or pinene.
- II. Nitrogen-containing compounds such as trimethylamine and methyl nitrate.
- III. Ketones; specifically acetone
- IV. Alcohols such as ethanol or methanol (isopropanol and propanol are common in cancer patients).

For sulphur compounds; dimethyl sulfide was reported by cancer studies while hydrogen sulfide appears indicative of liver disease.

Interpretation of volatile compounds from human breath is multifaceted and complex. Likely markers of cellular processes can be identified through knowledge of dominant metabolisms and considering systemic alterations and interactions. By considering markers of contrasting processes and pathophysiologies, the power of diagnosis will increase. Functional group targeting can help overcome variability within individuals and cohorts when looking for breath biomarkers of particular cellular functions. The 'breath-print' approach takes into account variability of biomarker metabolisms, conflicting comorbidities and physiological variations within individuals.

Application to COVID-19

A primary goal of this work is to provide contextual VOC targets, so that future research may target compounds with increased likelihood of diagnostic power. We may speculate upon how this work may relate to a critical topic in contemporary breath research: the diagnosis of viral lung infection, namely COVID-19. At this junction, with the limited published data available, we consider the underlying processes involved and compare this with research into other infections of the lung.

Several studies have explored whether COVID diagnosis via breath, using sensors and enose approaches [32,35–37], is plausible. Currently, 5 studies have been published which a) fit the criteria for inclusion in this article and ii) identify specific

VOCs as candidates for diagnosis. Substantial variability in COVID status exists within the studies undertaken, most notably, age of patients and disease severity at the point of breath collection [31,33,34,75,170]. Severity of disease influences VOCs seen in the breath (as shown in table 1). Severely ill patients, including those presenting with Acute Respiratory Distress syndrome (ARDs), will have impaired VOC diffusion into the lung space, due to the presence of fluid in the lungs. Furthermore, they may present with a range of disease complications outside of pulmonary ailments [171,172].

Published reports that fulfil our selection criteria report COVID representative functional groups as: aldehydes (notably C7 and over), carboxylic acids, oxygenated species, monoterpenes and halocarbons [31,33,34,75,170]. With an awareness of the limitations outlined and the large variability in collection and analysis methodology, we compared the functional outcomes from these papers against cancer and all other studies grouped (supplementary figure 3). COVID-19 revealed clear separation from cancer studies and sat within 'other' grouped studies with explained variance of 34.7% for PC1 and PC2 combined. One outlier for the COVID-19 group was identified as Ruszkiewicz *et al* 2020 due predominantly to the lack of hydrocarbons detected.

Pathophysiology of COVID-19 infections includes inflammatory response, characterised by oxidative stress (table 1) which has been linked to aldehydes and hydrocarbons [103]. Aldehydes are present in all 5 COVID-19 breath studies presented here and hydrocarbons are present in 4 studies. In comparison, studies investigating Influenza, a virus of the lung, revealed increased hydrocarbons in patients' breath following Influenza A vaccination [173] and pigs infected with Influenza revealed aldehydes in their breath [174].

As the volume of research around viral pathogens and volatile profiles grows, targets specific to pathogens will increase and the application of targets for early diagnosis aside from those targets linked to secondary and tertiary effects of infection will aid early application. We have demonstrated that researchers can consider targets from different functional groups and varying disease states

Conclusion

In conclusion, while mechanistic studies continue to be reported, and collections of cellular VOCs compiled [8] and contrasted with human breath databases [6,95,175,176], we contend that further information can be gained from comparing and contrasting breath profiles already reported within targeted metabolic and physiological contexts and that this approach will help inform further research. We have demonstrated that commonality exists in a suite of volatiles present in the breath of patients across a range of diseases and that these volatiles can also separate disease groups.

Acknowledgements

This work has been made possible through the white rose mechanistic biology doctoral training program. Supported by the Biotechnology and Biological Science Research Council (BBSRC) BB/M011151/1.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Accepted Manuscript

1. Kwak J, Preti G. Volatile disease biomarkers in breath: a critique. *Curr Pharm Biotechnol.* 2011;12: 1067–1074.
2. Blanchet L, Smolinska A, Baranska A, Tigchelaar E, Swertz M, Zhernakova A, et al. Factors that influence the volatile organic compound content in human breath. *J Breath Res.* 2017;11: 016013.
3. Das S, Pal M. Review—Non-Invasive Monitoring of Human Health by Exhaled Breath Analysis: A Comprehensive Review. *J Electrochem Soc.* 2020;167: 037562.
4. Amann A, Costello B de L, Miekisch W, Schubert J, Buszewski B, Pleil J, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J Breath Res.* 2014;8: 034001.
5. Drabińska N, Flynn C, Ratcliffe N, Belluomo I, Myridakis A, Gould O, et al. A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *J Breath Res.* 2021;15. doi:10.1088/1752-7163/abf1d0
6. Jia Z, Patra A, Kutty VK, Venkatesan T. Critical Review of Volatile Organic Compound Analysis in Breath and In Vitro Cell Culture for Detection of Lung Cancer. *Metabolites.* 2019;9. doi:10.3390/metabo9030052
7. Gasparri R, Santonico M, Valentini C, Sedda G, Borri A, Petrella F, et al. Volatile signature for the early diagnosis of lung cancer. *J Breath Res.* 2016;10: 016007.
8. Filipiak W, Mochalski P, Filipiak A, Ager C, Cumeras R, Davis CE, et al. A Compendium of Volatile Organic Compounds (VOCs) Released By Human Cell Lines. *Curr Med Chem.* 2016;23: 2112–2131.
9. Holz O, Waschki B, Schuchardt S, Watz H, Fuchs O, Von Mutius E, et al. Breath volatile organic compound (VOC) patterns in adult asthma patients of the ALLIANCE cohort. *Eur Respir J.* 2019;54. doi:10.1183/13993003.congress-2019.PA1711
10. Smolinska A, Klaassen EMM, Dallinga JW, van de Kant KDG, Jobsis Q, Moonen EJC, et al. Profiling of Volatile Organic Compounds in Exhaled Breath As a Strategy to Find Early Predictive Signatures of Asthma in Children. *PLoS ONE.* 2014. p. e95668. doi:10.1371/journal.pone.0095668
11. Gaida A, Holz O, Nell C, Schuchardt S, Lavae-Mokhtari B, Kruse L, et al. A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD. *J Breath Res.* 2016;10: 026006.
12. Santini G, Mores N, Shohreh R, Valente S, Dabrowska M, Trové A, et al. Exhaled and non-exhaled non-invasive markers for assessment of respiratory inflammation in patients with stable COPD and healthy smokers. *J Breath Res.* 2016;10: 017102.
13. van Mastriigt E, Reyes-Reyes A, Brand K, Bhattacharya N, Urbach HP, Stubbs AP, et al. Exhaled breath profiling using broadband quantum cascade laser-based spectroscopy in healthy children and children with asthma and cystic fibrosis. *J Breath Res.* 2016;10: 026003.
14. Paredi P, Kharitonov SA, Leak D, Shah PL, Cramer D, Hodson ME, et al. Exhaled Ethane Is Elevated in Cystic Fibrosis and Correlates with Carbon Monoxide Levels and Airway Obstruction. *American Journal of Respiratory and Critical Care Medicine.* 2000. pp. 1247–1251. doi:10.1164/ajrccm.161.4.9906122

15. Saktiawati AMI, Putera DD, Setyawan A, Mahendradhata Y, van der Werf TS. Diagnosis of tuberculosis through breath test: A systematic review. *EBioMedicine*. 2019;46: 202–214.
16. Beccaria M, Mellors TR, Petion JS, Rees CA, Nasir M, Systrom HK, et al. Preliminary investigation of human exhaled breath for tuberculosis diagnosis by multidimensional gas chromatography – Time of flight mass spectrometry and machine learning. *Journal of Chromatography B*. 2018. pp. 46–50. doi:10.1016/j.jchromb.2018.01.004
17. Janfaza S, Khorsand B, Nikkhah M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. *Biol Methods Protoc*. 2019;4: bpz014.
18. Bykova AA, Malinovskaya LK, Chomakhidze PS, Trushina OV, Shaltaeva YR, Belyakov VV, et al. [Exhaled Breath Analysis in Diagnostics of Cardiovascular Diseases]. *Kardiologija*. 2019;59: 61–67.
19. De Vincentis A, Vespasiani-Gentilucci U, Sabatini A, Antonelli-Incalzi R, Picardi A. Exhaled breath analysis in hepatology: State-of-the-art and perspectives. *World J Gastroenterol*. 2019;25: 4043–4050.
20. Bannaga AS, Farrugia A, Arasaradnam RP. Diagnosing Inflammatory bowel disease using noninvasive applications of volatile organic compounds: a systematic review. *Expert Rev Gastroenterol Hepatol*. 2019;13: 1113–1122.
21. Tiele A, Wicaksono A, Daulton E, Ifeachor E, Eyre V, Clarke S, et al. Breath-based non-invasive diagnosis of Alzheimer's disease: a pilot study. *J Breath Res*. 2020;14: 026003.
22. Tisch U, Schlesinger I, Ionescu R, Nassar M, Axelrod N, Robertman D, et al. Detection of Alzheimer's and Parkinson's disease from exhaled breath using nanomaterial-based sensors. *Nanomedicine* . 2013;8: 43–56.
23. Das S, Pal S, Mitra M. Significance of Exhaled Breath Test in Clinical Diagnosis: A Special Focus on the Detection of Diabetes Mellitus. *J Med Biol Eng*. 2016;36: 605–624.
24. Berna AZ, Odom John AR. Breath Metabolites to Diagnose Infection. *Clin Chem*. 2021;68: 43–51.
25. Ghosh C, Leon A, Koshy S, Aloum O, Al-Jabawi Y, Ismail N, et al. Breath-Based Diagnosis of Infectious Diseases: A Review of the Current Landscape. *Clin Lab Med*. 2021;41: 185–202.
26. Phillips M, Cataneo RN, Condos R, Ring Erickson GA, Greenberg J, La Bombardi V, et al. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* . 2007;87: 44–52.
27. Scott-Thomas AJ, Syhre M, Pattermore PK, Epton M, Laing R, Pearson J, et al. 2-Aminoacetophenone as a potential breath biomarker for *Pseudomonas aeruginosa* in the cystic fibrosis lung. *BMC Pulm Med*. 2010;10: 56.
28. Shestivska V, Nemeč A, Dřevínek P, Sovová K, Dryahina K, Spaněl P. Quantification of methyl thiocyanate in the headspace of *Pseudomonas aeruginosa* cultures and in the breath of cystic fibrosis patients by selected ion flow tube mass spectrometry. *Rapid Commun Mass Spectrom*. 2011;25: 2459–2467.
29. Mendel J, Frank K, Edlin L, Hall K, Webb D, Mills J, et al. Preliminary accuracy of COVID-19 odor detection by canines and HS-SPME-GC-MS using exhaled breath

- 1
2
3 samples. *Forensic Sci Int Synerg.* 2021;3: 100155.
4
- 5 30. Jendryn P, Schulz C, Twele F, Meller S, von Köckritz-Blickwede M, Osterhaus ADME, et
6 al. Scent dog identification of samples from COVID-19 patients - a pilot study. *BMC*
7 *Infect Dis.* 2020;20: 536.
8
- 9 31. Ruszkiewicz DM, Sanders D, O'Brien R, Hempel F, Reed MJ, Riepe AC, et al.
10 Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility
11 spectrometry - a feasibility study. *EClinicalMedicine.* 2020;29: 100609.
12
- 13 32. Chen H, Qi X, Zhang L, Li X, Ma J, Zhang C, et al. COVID-19 screening using breath-
14 borne volatile organic compounds. *J Breath Res.* 2021;15. doi:10.1088/1752-
15 7163/ac2e57
16
- 17 33. Grassin-Delyle S, Roquencourt C, Moine P, Saffroy G, Carn S, Heming N, et al.
18 Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study.
19 *EBioMedicine.* 2021;63: 103154.
20
- 21 34. Berna AZ, Akaho EH, Harris RM, Congdon M, Korn E, Neher S, et al. Reproducible
22 Breath Metabolite Changes in Children with SARS-CoV-2 Infection. *ACS Infect Dis.*
23 2021;7: 2596–2603.
24
- 25 35. Wintjens AGWE, Hintzen KFH, Engelen SME, Lubbers T, Savelkoul PHM, Wesseling G,
26 et al. Applying the electronic nose for pre-operative SARS-CoV-2 screening. *Surg*
27 *Endosc.* 2021;35: 6671–6678.
28
- 29 36. de Vries R, Vigeveno RM, Mulder S, Farzan N, Vintges DR, Goeman JJ, et al. Ruling
30 out SARS-CoV-2 infection using exhaled breath analysis by electronic nose in a public
31 health setting. *bioRxiv. medRxiv;* 2021. doi:10.1101/2021.02.14.21251712
32
- 33 37. Shan B, Broza YY, Li W, Wang Y, Wu S, Liu Z, et al. Multiplexed Nanomaterial-Based
34 Sensor Array for Detection of COVID-19 in Exhaled Breath. *ACS Nano.* 2020;14:
35 12125–12132.
36
- 37 38. Subali AD, Wiyono L, Yusuf M, Zaky MFA. The potential of volatile organic compounds-
38 based breath analysis for COVID-19 screening: a systematic review & meta-analysis.
39 *Diagn Microbiol Infect Dis.* 2022;102: 115589.
40
- 41 39. Miettinen OS, Steurer J, Hofman A. The Bayes' Theorem Framework for Diagnostic
42 Research. *Clinical Research Transformed.* 2019. pp. 109–114. doi:10.1007/978-3-030-
43 06176-0_13
44
- 45 40. O'Hara ME, Fernández Del Río R, Holt A, Pemberton P, Shah T, Whitehouse T, et al.
46 Limonene in exhaled breath is elevated in hepatic encephalopathy. *J Breath Res.*
47 2016;10: 046010.
48
- 49 41. Molina-Molina E, Shanmugam H, Di Palo D, Grattagliano I, Portincasa P. Exploring
50 Liver Mitochondrial Function by ¹³C-Stable Isotope Breath Tests: Implications in Clinical
51 Biochemistry. *Methods Mol Biol.* 2021;2310: 179–199.
52
- 53 42. Gaude E, Nakhleh MK, Patassini S, Boschmans J, Allsworth M, Boyle B, et al. Targeted
54 breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-
55 specific probes. *J Breath Res.* 2019;13: 032001.
56
- 57 43. Sangnes DA, Søfteland E, Teigland T, Dimcevski G. Comparing radiopaque markers
58 and ¹³C-labelled breath test in diabetic gastroparesis diagnostics. *Clin Exp*
59
60

- 1
2
3 Gastroenterol. 2019;12: 193–201.
4
- 5 44. Savarino V, Vigneri S, Celle G. The ¹³C urea breath test in the diagnosis of
6 *Helicobacter pylori* infection. *Gut*. 1999;45 Suppl 1: 118–22.
7
- 8 45. Modak AS. Stable isotope breath tests in clinical medicine: a review. *J Breath Res*.
9 2007;1: 014003.
10
- 11 46. Born P, Zech J, Lehn H, Classen M, Lorenz R. Colonic bacterial activity determines the
12 symptoms in people with fructose-malabsorption. *Hepatology*. 1995;42:
13 778–785.
14
- 15 47. Helwig U, Koch AK, Koppka N, Holtmann S, Langhorst J. The Predictive Value of the
16 Hydrogen Breath Test in the Diagnosis of Fructose Malabsorption. *Digestion*. 2019;99:
17 140–147.
18
- 19 48. Westhoff M, Rickermann M, Litterst P, Baumbach JI. Exogenous factors of influence on
20 exhaled breath analysis by ion-mobility spectrometry (MCC/IMS). *Int J Ion Mobil*
21 *Spectrom*. 2019;22: 59–69.
22
- 23 49. Capone S, Tufariello M, Forleo A, Longo V, Giampetruzzi L, Radogna AV, et al.
24 Chromatographic analysis of VOC patterns in exhaled breath from smokers and
25 nonsmokers. *Biomed Chromatogr*. 2018;32. doi:10.1002/bmc.4132
26
- 27 50. Wallace MAG, Ariel Geer Wallace M, Pleil JD, Oliver KD, Whitaker DA, Mentese S, et
28 al. Non-targeted GC/MS analysis of exhaled breath samples: Exploring human
29 biomarkers of exogenous exposure and endogenous response from professional
30 firefighting activity. *Journal of Toxicology and Environmental Health, Part A*. 2019. pp.
31 244–260. doi:10.1080/15287394.2019.1587901
32
- 33 51. Li AJ, Pal VK, Kannan K. A review of environmental occurrence, toxicity,
34 biotransformation and biomonitoring of volatile organic compounds. *Environmental*
35 *Chemistry and Ecotoxicology*. 2021;3: 91–116.
36
- 37 52. Proctor LM. The Human Microbiome Project in 2011 and beyond. *Cell Host Microbe*.
38 2011;10: 287–291.
39
- 40 53. Bos LDJ, Meinardi S, Blake D, Whiteson K. Bacteria in the airways of patients with
41 cystic fibrosis are genetically capable of producing VOCs in breath. *J Breath Res*.
42 2016;10: 047103.
43
- 44 54. Sohrabi M, Zhang L, Zhang K, Ahmetagic A, Wei MQ. Volatile organic compounds as
45 novel markers for the detection of bacterial infections. *Clinical Microbiology: Open*
46 *Access*. 2014. Available:
47 https://www.researchgate.net/profile/Li_Zhang27/publication/264899258_Volatile_Organic_Compounds_as_Novel_Markers_for_the_Detection_of_Bacterial_Infections/links/53f578610cf2fceacc6f4640.pdf
48
49
50
- 51 55. Seesaard T, Thippakorn C, Kerdcharoen T, Kladsomboon S. A hybrid electronic nose
52 system for discrimination of pathogenic bacterial volatile compounds. *Anal Methods*.
53 2020;12: 5671–5683.
54
- 55 56. Dorokhov YL, Shindyapina AV, Sheshukova EV, Komarova TV. Metabolic methanol:
56 molecular pathways and physiological roles. *Physiol Rev*. 2015;95: 603–644.
57
- 58 57. Bodelier AGL, Smolinska A, Baranska A, Dallinga JW, Mujagic Z, Vanhees K, et al.
59
60

- 1
2
3 Volatile Organic Compounds in Exhaled Air as Novel Marker for Disease Activity in
4 Crohn's Disease: A Metabolomic Approach. *Inflamm Bowel Dis.* 2015;21: 1776–1785.
5
- 6 58. Kramer C, Mochalski P, Unterkofler K, Agapiou A, Ruzsanyi V, Liedl KR. Prediction of
7 blood:air and fat:air partition coefficients of volatile organic compounds for the
8 interpretation of data in breath gas analysis. *J Breath Res.* 2016;10: 017103.
9
- 10 59. King J, Unterkofler K, Teschl G, Teschl S, Mochalski P, Koç H, et al. A modeling-based
11 evaluation of isothermal rebreathing for breath gas analyses of highly soluble volatile
12 organic compounds. *J Breath Res.* 2012;6: 016005.
13
- 14 60. Statheropoulos M, Agapiou A, Georgiadou A. Analysis of expired air of fasting male
15 monks at Mount Athos. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2006;832:
16 274–279.
17
- 18 61. Galassetti PR, Novak B, Nemet D, Rose-Gottron C, Cooper DM, Meinardi S, et al.
19 Breath Ethanol and Acetone as Indicators of Serum Glucose Levels: An Initial Report.
20 *Diabetes Technology & Therapeutics.* 2005. pp. 115–123. doi:10.1089/dia.2005.7.115
21
- 22 62. Strambi A, De Milito A. Contribution of pH Alterations to the Tumor Microenvironment.
23 In: Mazurek S, Shoshan M, editors. *Tumor Cell Metabolism: Pathways, Regulation and*
24 *Biology.* Vienna: Springer Vienna; 2015. pp. 173–196.
25
- 26 63. Petrova V, Annicchiarico-Petruzzelli M, Melino G, Amelio I. The hypoxic tumour
27 microenvironment. *Oncogenesis.* 2018;7: 10.
28
- 29 64. Leslie TK, James AD, Zaccagna F, Grist JT, Deen S, Kennerley A, et al. Sodium
30 homeostasis in the tumour microenvironment. *Biochim Biophys Acta Rev Cancer.*
31 2019;1872: 188304.
32
- 33 65. Issitt T, Bosseboeuf E, De Winter N, Dufton N, Gestri G, Senatore V, et al. Neuropilin-1
34 Controls Endothelial Homeostasis by Regulating Mitochondrial Function and Iron-
35 Dependent Oxidative Stress. *iScience.* 2019;11: 205–223.
36
- 37 66. He S, Sharpless NE. Senescence in Health and Disease. *Cell.* 2017. pp. 1000–1011.
38 doi:10.1016/j.cell.2017.05.015
39
- 40 67. Gaschler MM, Stockwell BR. Lipid peroxidation in cell death. *Biochemical and*
41 *Biophysical Research Communications.* 2017. pp. 419–425.
42 doi:10.1016/j.bbrc.2016.10.086
43
- 44 68. Phillips M, Greenberg J, Awad J. Metabolic and environmental origins of volatile organic
45 compounds in breath. *J Clin Pathol.* 1994;47: 1052–1053.
46
- 47 69. Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic
48 compounds of lung cancer and possible biochemical pathways. *Chem Rev.* 2012;112:
49 5949–5966.
50
- 51 70. Massey KA, Nicolaou A. Lipidomics of polyunsaturated-fatty-acid-derived oxygenated
52 metabolites. *Biochem Soc Trans.* 2011;39: 1240–1246.
53
- 54 71. Esterbauer H, Schaur RJ, Zollner H. Chemistry and biochemistry of 4-hydroxynonenal,
55 malonaldehyde and related aldehydes. *Free Radic Biol Med.* 1991;11: 81–128.
56
- 57 72. Kim K-H, Jahan SA, Kabir E. A review of breath analysis for diagnosis of human health.
58 *Trends Analyt Chem.* 2012;33: 1–8.
59
60

- 1
- 2
- 3 73. Beauchamp J, Herbig J, Gutmann R, Hansel A. On the use of Tedlar® bags for breath-
- 4 gas sampling and analysis. *J Breath Res.* 2008;2: 046001.
- 5
- 6 74. Ghimenti S, Lomonaco T, Bellagambi FG, Tabucchi S, Onor M, Trivella MG, et al.
- 7 Comparison of sampling bags for the analysis of volatile organic compounds in breath. *J*
- 8 *Breath Res.* 2015;9: 047110.
- 9
- 10 75. Liangou A, Tasoglou A, Huber HJ, Wistrom C, Brody K, Menon PG, et al. A method for
- 11 the identification of COVID-19 biomarkers in human breath using Proton Transfer
- 12 Reaction Time-of-Flight Mass Spectrometry. *EClinicalMedicine.* 2021;42: 101207.
- 13
- 14 76. Ibrahim W, Carr L, Cordell R, Wilde MJ, Salman D, Monks PS, et al. Breathomics for the
- 15 clinician: the use of volatile organic compounds in respiratory diseases. *Thorax.*
- 16 2021;76: 514–521.
- 17
- 18 77. Chen X, Cao M, Hao Y, Li Y, Wang P, Ying K, et al. A Non-invasive detection of lung
- 19 cancer combined virtual gas sensors array with imaging recognition technique. *Conf*
- 20 *Proc IEEE Eng Med Biol Soc.* 2005;2005: 5873–5876.
- 21
- 22 78. Ma H, Li X, Chen J, Wang H, Cheng T, Chen K, et al. Analysis of human breath
- 23 samples of lung cancer patients and healthy controls with solid-phase microextraction
- 24 (SPME) and flow-modulated comprehensive two-dimensional gas chromatography (GC
- 25 × GC). *Analytical Methods.* 2014. p. 6841. doi:10.1039/c4ay01220h
- 26
- 27 79. Ulanowska A, Kowalkowski T, Trawińska E, Buszewski B. The application of statistical
- 28 methods using VOCs to identify patients with lung cancer. *J Breath Res.* 2011;5:
- 29 046008.
- 30
- 31 80. Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, et al. Exhaled volatile
- 32 organic compounds in patients with non-small cell lung cancer: cross sectional and
- 33 nested short-term follow-up study. *Respir Res.* 2005;6: 71.
- 34
- 35 81. Wehinger A, Schmid A, Mechtcheriakov S, Ledochowski M, Grabmer C, Gastl GA, et al.
- 36 Lung cancer detection by proton transfer reaction mass-spectrometric analysis of
- 37 human breath gas. *Int J Mass Spectrom.* 2007;265: 49–59.
- 38
- 39 82. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, et al. Noninvasive
- 40 detection of lung cancer by analysis of exhaled breath. *BMC Cancer.* 2009;9: 348.
- 41
- 42 83. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and Methodologic Challenges
- 43 of Volatile Organic Compound-Based Exhaled Breath Tests for Cancer Diagnosis: A
- 44 Systematic Review and Meta-analysis. *JAMA Oncol.* 2019;5: e182815.
- 45
- 46 84. Van Malderen K, De Winter BY, De Man JG, De Schepper HU, Lamote K. Volatomics in
- 47 inflammatory bowel disease and irritable bowel syndrome. *EBioMedicine.* 2020;54:
- 48 102725.
- 49
- 50 85. Markar SR, Wiggins T, Kumar S, Hanna GB. Exhaled breath analysis for the diagnosis
- 51 and assessment of endoluminal gastrointestinal diseases. *J Clin Gastroenterol.*
- 52 2015;49: 1–8.
- 53
- 54 86. Issitt, TJ, Redeker K. Volatile compounds in human breath: a review and meta-analysis.
- 55 PROSPERO 2021 CRD42021234660. 2021. Available:
- 56 https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021234660
- 57
- 58
- 59 87. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The
- 60

- 1
2
3 PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*.
4 2021;372: n71.
5
- 6 88. Christiansen A, Davidsen JR, Titlestad I, Vestbo J, Baumbach J. A systematic review of
7 breath analysis and detection of volatile organic compounds in COPD. *J Breath Res*.
8 2016;10: 034002.
9
- 10 89. Hicks LC, Huang J, Kumar S, Powles ST, Orchard TR, Hanna GB, et al. Analysis of
11 Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot
12 Study. *J Crohns Colitis*. 2015;9: 731–737.
13
- 14 90. Khalid TY, Costello BDL, Ewen R, White P, Stevens S, Gordon F, et al. Breath volatile
15 analysis from patients diagnosed with harmful drinking, cirrhosis and hepatic
16 encephalopathy: a pilot study. *Metabolomics*. 2013;9: 938–948.
17
- 18 91. Patel N, Alkhoury N, Eng K, Cikach F, Mahajan L, Yan C, et al. Metabolomic analysis of
19 breath volatile organic compounds reveals unique breathprints in children with
20 inflammatory bowel disease: a pilot study. *Aliment Pharmacol Ther*. 2014;40: 498–507.
21
- 22 92. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile
23 organic compounds in breath as markers of lung cancer: a cross-sectional study.
24 *Lancet*. 1999;353: 1930–1933.
25
- 26 93. Sahota AS, Gowda R, Arasaradnam RP, Daulton E, Savage RS, Skinner JR, et al. A
27 simple breath test for tuberculosis using ion mobility: A pilot study. *Tuberculosis* .
28 2016;99: 143–146.
29
- 30 94. Alkhoury N, Singh T, Alsabbagh E, Guirguis J, Chami T, Hanouneh I, et al. Isoprene in
31 the Exhaled Breath is a Novel Biomarker for Advanced Fibrosis in Patients with Chronic
32 Liver Disease: A Pilot Study. *Clin Transl Gastroenterol*. 2015;6: e112.
33
- 34 95. Kuo T-C, Tan C-E, Wang S-Y, Lin OA, Su B-H, Hsu M-T, et al. Human Breathomics
35 Database. *Database* . 2020;2020. doi:10.1093/database/baz139
36
- 37 96. Tejero Rioseras A, Singh KD, Nowak N, Gaugg MT, Bruderer T, Zenobi R, et al. Real-
38 Time Monitoring of Tricarboxylic Acid Metabolites in Exhaled Breath. *Anal Chem*.
39 2018;90: 6453–6460.
40
- 41 97. Tiele A, Wicaksono A, Kansara J, Arasaradnam RP, Covington JA. Breath Analysis
42 Using eNose and Ion Mobility Technology to Diagnose Inflammatory Bowel Disease-A
43 Pilot Study. *Biosensors*. 2019;9. doi:10.3390/bios9020055
44
- 45 98. Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, et al.
46 Review of recent developments in GC–MS approaches to metabolomics-based
47 research. *Metabolomics*. 2018;14: 152.
48
- 49 99. Hanouneh IA, Zein NN, Cikach F, Dababneh L, Grove D, Alkhoury N, et al. The
50 breathprints in patients with liver disease identify novel breath biomarkers in alcoholic
51 hepatitis. *Clin Gastroenterol Hepatol*. 2014;12: 516–523.
52
- 53 100. Hanson JR. *Functional Group Chemistry*. Royal Society of Chemistry; 2001.
54
- 55 101. King J, Koc H, Unterkofler K, Mochalski P, Kupferthaler A, Teschl G, et al. Physiological
56 modeling of isoprene dynamics in exhaled breath. *J Theor Biol*. 2010;267: 626–637.
57
- 58 102. Phillips M, Cataneo RN, Greenberg J, Gunawardena R, Naidu A, Rahbari-Oskoui F.
59 Effect of age on the breath methylated alkane contour, a display of apparent new
60

- 1
2
3 markers of oxidative stress. *J Lab Clin Med.* 2000;136: 243–249.
4
- 5 103. Ratcliffe N, Wieczorek T, Drabińska N, Gould O, Osborne A, De Lacy Costello B. A
6 mechanistic study and review of volatile products from peroxidation of unsaturated fatty
7 acids: an aid to understanding the origins of volatile organic compounds from the human
8 body. *J Breath Res.* 2020;14: 034001.
9
- 10 104. Netzer M, Millonig G, Osl M, Pfeifer B, Praun S, Villinger J, et al. A new ensemble-
11 based algorithm for identifying breath gas marker candidates in liver disease using ion
12 molecule reaction mass spectrometry. *Bioinformatics.* 2009;25: 941–947.
13
- 14 105. Miekisch W, Herbig J, Schubert JK. Data interpretation in breath biomarker research:
15 pitfalls and directions. *J Breath Res.* 2012;6: 036007.
16
- 17 106. Miekisch W, Schubert JK, Noeldge-Schomburg GFE. Diagnostic potential of breath
18 analysis--focus on volatile organic compounds. *Clin Chim Acta.* 2004;347: 25–39.
19
- 20 107. Mazzone PJ. Analysis of volatile organic compounds in the exhaled breath for the
21 diagnosis of lung cancer. *J Thorac Oncol.* 2008;3: 774–780.
22
- 23 108. Amann A, Corradi M, Mazzone P, Mutti A. Lung cancer biomarkers in exhaled breath.
24 *Expert Rev Mol Diagn.* 2011;11: 207–217.
25
- 26 109. Zhou M, Liu Y, Duan Y. Breath biomarkers in diagnosis of pulmonary diseases. *Clin*
27 *Chim Acta.* 2012;413: 1770–1780.
28
- 29 110. Minh TDC, Blake DR, Galassetti PR. The clinical potential of exhaled breath analysis for
30 diabetes mellitus. *Diabetes Res Clin Pract.* 2012;97: 195–205.
31
- 32 111. Riely CA, Cohen G, Lieberman M. Ethane evolution: a new index of lipid peroxidation.
33 *Science.* 1974;183: 208–210.
34
- 35 112. Negre-Salvayre A, Coatrieux C, Ingueneau C, Salvayre R. Advanced lipid peroxidation
36 end products in oxidative damage to proteins. Potential role in diseases and therapeutic
37 prospects for the inhibitors. *Br J Pharmacol.* 2008;153: 6–20.
38
- 39 113. Sobotka PA, Gupta DK, Lansky DM, Costanzo MR, Zarling EJ. Breath pentane is a
40 marker of acute cardiac allograft rejection. *J Heart Lung Transplant.* 1994;13: 224–229.
41
- 42 114. Kneepkens CM, Lepage G, Roy CC. The potential of the hydrocarbon breath test as a
43 measure of lipid peroxidation. *Free Radic Biol Med.* 1994;17: 127–160.
44
- 45 115. Müller A, Sies H. Role of alcohol dehydrogenase activity and the acetaldehyde in
46 ethanol- induced ethane and pentane production by isolated perfused rat liver. *Biochem*
47 *J.* 1982;206: 153–156.
48
- 49 116. Bornhorst JA, Mbughuni MM. Chapter 3 - Alcohol Biomarkers: Clinical Issues and
50 Analytical Methods. In: Dasgupta A, editor. *Critical Issues in Alcohol and Drugs of*
51 *Abuse Testing (Second Edition).* Academic Press; 2019. pp. 25–42.
52
- 53 117. Davis PL, Dal Cortivo LA, Maturo J. Endogenous isopropanol: forensic and biochemical
54 implications. *J Anal Toxicol.* 1984;8: 209–212.
55
- 56 118. Dadamio J, Van den Velde S, Laleman W, Van Hee P, Coucke W, Nevens F, et al.
57 Breath biomarkers of liver cirrhosis. *J Chromatogr B Analyt Technol Biomed Life Sci.*
58 2012;905: 17–22.
59
60

119. Pedersen KJ. THE KETONIC DECOMPOSITION OF BETA-KETO CARBOXYLIC ACIDS. *J Am Chem Soc.* 1929;51: 2098–2107.
120. López-Soriano FJ, Alemany M, Argilés JM. Rat acetoacetic acid decarboxylase inhibition by acetone. *Int J Biochem.* 1985;17: 1271–1273.
121. Ruzsányi V, Péter Kalapos M. Breath acetone as a potential marker in clinical practice. *J Breath Res.* 2017;11: 024002.
122. Nordmann R, Ribiere C, Rouach H, Beauge F, Giudicelli Y, Nordmann J. Metabolic pathways involved in the oxidation of isopropanol into acetone by the intact rat. *Life Sci.* 1973;13: 919–932.
123. Erhart S, Amann A, Haberlandt E, Edlinger G, Schmid A, Filipiak W, et al. 3-Heptanone as a potential new marker for valproic acid therapy. *J Breath Res.* 2009;3: 016004.
124. Ye M, Chien P-J, Toma K, Arakawa T, Mitsubayashi K. An acetone bio-sniffer (gas phase biosensor) enabling assessment of lipid metabolism from exhaled breath. *Biosens Bioelectron.* 2015;73: 208–213.
125. Shahidi F. Headspace volatile aldehydes as indicators of lipid oxidation in foods. *Adv Exp Med Biol.* 2001;488: 113–123.
126. Ayala A, Muñoz MF, Argüelles S. Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxid Med Cell Longev.* 2014;2014: 360438.
127. Wickramasinghe SN, Bond AN, Sloviter HA, Saunders JE. Metabolism of ethanol by human bone marrow cells. *Acta Haematol.* 1981;66: 238–243.
128. Mellick GD. CYP450, genetics and Parkinson's disease: genex environment interactions hold the key. *Parkinson's Disease and Related Disorders.* Springer; 2006. pp. 159–165.
129. Larkin DR. The role of catalysts in the air oxidation of aliphatic aldehydes. *J Org Chem.* 1990;55: 1563–1568.
130. Callol-Sanchez L, Munoz-Lucas MA, Gomez-Martin O, Maldonado-Sanz JA, Civera-Tejuca C, Gutierrez-Ortega C, et al. Observation of nonanoic acid and aldehydes in exhaled breath of patients with lung cancer. *J Breath Res.* 2017;11: 026004.
131. Jareño-Esteban JJ, Muñoz-Lucas MÁ, Gómez-Martín Ó, Utrilla-Trigo S, Gutiérrez-Ortega C, Aguilar-Ros A, et al. Study of 5 Volatile Organic Compounds in Exhaled Breath in Chronic Obstructive Pulmonary Disease. *Archivos de Bronconeumología (English Edition).* 2017;53: 251–256.
132. Dryahina K, Smith D, Bortlík M, Machková N, Lukáš M, Španěl P. Pentane and other volatile organic compounds, including carboxylic acids, in the exhaled breath of patients with Crohn's disease and ulcerative colitis. *J Breath Res.* 2017;12: 016002.
133. Fukami T, Yokoi T. The emerging role of human esterases. *Drug Metab Pharmacokinet.* 2012;27: 466–477.
134. Ye W, Zhang Y, He M, Zhu C, Feng X-P. Relationship of tongue coating microbiome on volatile sulfur compounds in healthy and halitosis adults. *J Breath Res.* 2019;14: 016005.
135. Snel J, Burgering M, Smit B, Noordman W, Tangerman A, Winkel EG, et al. Volatile

- 1
2
3 sulphur compounds in morning breath of human volunteers. *Arch Oral Biol.* 2011;56:
4 29–34.
5
- 6 136. Trefz P, Koehler H, Klepik K, Moebius P, Reinhold P, Schubert JK, et al. Volatile
7 emissions from *Mycobacterium avium* subsp. *paratuberculosis* mirror bacterial growth
8 and enable distinction of different strains. *PLoS One.* 2013;8: e76868.
9
- 10 137. Shimamoto C, Hirata I, Katsu K. Breath and blood ammonia in liver cirrhosis.
11 *Hepatogastroenterology.* 2000;47: 443–445.
12
- 13 138. Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and
14 implementation of breath volatile cancer markers. *Chem Soc Rev.* 2014;43: 1423–1449.
15
- 16 139. Guengerich FP, Peter Guengerich F, Shimada T. Oxidation of toxic and carcinogenic
17 chemicals by human cytochrome P-450 enzymes. *Chemical Research in Toxicology.*
18 1991. pp. 391–407. doi:10.1021/tx00022a001
19
- 20 140. Buszewski B, Ulanowska A, Ligor T, Denderz N, Amann A. Analysis of exhaled breath
21 from smokers, passive smokers and non-smokers by solid-phase microextraction gas
22 chromatography/mass spectrometry. *Biomedical Chromatography.* 2009. pp. 551–556.
23 doi:10.1002/bmc.1141
24
- 25 141. Rudnicka J, Kowalkowski T, Ligor T, Buszewski B. Determination of volatile organic
26 compounds as biomarkers of lung cancer by SPME-GC-TOF/MS and chemometrics. *J*
27 *Chromatogr B Analyt Technol Biomed Life Sci.* 2011;879: 3360–3366.
28
- 29 142. Thriumani R, Zakaria A, Hashim YZH-Y, Jeffree AI, Helmy KM, Kamarudin LM, et al. A
30 study on volatile organic compounds emitted by in-vitro lung cancer cultured cells using
31 gas sensor array and SPME-GCMS. *BMC Cancer.* 2018;18: 362.
32
- 33 143. Silva CL, Perestrelo R, Silva P, Tomás H, Câmara JS. Volatile metabolomic signature of
34 human breast cancer cell lines. *Sci Rep.* 2017;7: 43969.
35
- 36 144. Hanai Y, Shimono K, Oka H, Baba Y, Yamazaki K, Beauchamp GK. Analysis of volatile
37 organic compounds released from human lung cancer cells and from the urine of tumor-
38 bearing mice. *Cancer Cell Int.* 2012;12: 7.
39
- 40 145. Jia Z, Zhang H, Ong CN, Patra A, Lu Y, Lim CT, et al. Detection of Lung Cancer:
41 Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell
42 Lines of Different Histological Origins. *ACS Omega.* 2018;3: 5131–5140.
43
- 44 146. Serasanambati M, Broza YY, Marmur A, Haick H. Profiling Single Cancer Cells with
45 Volatolomics Approach. *iScience.* 2019;11: 178–188.
46
- 47 147. Peled N, Barash O, Tisch U, Ionescu R, Broza YY, Ilouze M, et al. Volatile fingerprints of
48 cancer specific genetic mutations. *Nanomedicine.* 2013;9: 758–766.
49
- 50 148. Kwak J, Gallagher M, Ozdener MH, Wysocki CJ, Goldsmith BR, Isamah A, et al. Volatile
51 biomarkers from human melanoma cells. *J Chromatogr B Analyt Technol Biomed Life*
52 *Sci.* 2013;931: 90–96.
53
- 54 149. Filipiak W, Sponring A, Filipiak A, Ager C, Schubert J, Miekisch W, et al. TD-GC-MS
55 analysis of volatile metabolites of human lung cancer and normal cells in vitro. *Cancer*
56 *Epidemiol Biomarkers Prev.* 2010;19: 182–195.
57
- 58 150. Filipiak W, Ruzsanyi V, Mochalski P, Filipiak A, Bajtarevic A, Ager C, et al. Dependence
59 of exhaled breath composition on exogenous factors, smoking habits and exposure to
60

- 1
2
3 air pollutants. *J Breath Res.* 2012;6: 036008.
4
- 5 151. Zoller O, Sager F, Reinhard H. Furan in food: headspace method and product survey.
6 *Food Addit Contam.* 2007;24 Suppl 1: 91–107.
7
- 8 152. Kushch I, Schwarz K, Schwentner L, Baumann B, Dzien A, Schmid A, et al. Compounds
9 enhanced in a mass spectrometric profile of smokers' exhaled breath versus non-
10 smokers as determined in a pilot study using PTR-MS. *J Breath Res.* 2008;2: 026002.
11
- 12 153. Fielding D, Hartel G, Pass D, Davis M, Brown M, Dent A, et al. Volatile organic
13 compound breath testing detects in-situ squamous cell carcinoma of bronchial and
14 laryngeal regions and shows distinct profiles of each tumour. *J Breath Res.* 2020;14:
15 046013.
16
- 17 154. Little LD, Carolan VA, Allen KE, Cole LM, Haywood-Small SL. Headspace analysis of
18 mesothelioma cell lines differentiates biphasic and epithelioid sub-types. *J Breath Res.*
19 2020;14: 046011.
20
- 21 155. Bhattacharyya D, Kumar P, Mohanty SK, Smith YR, Misra M. Detection of Four Distinct
22 Volatile Indicators of Colorectal Cancer using Functionalized Titania Nanotubular
23 Arrays. *Sensors* . 2017;17. doi:10.3390/s17081795
24
- 25 156. Handa H, Usuba A, Maddula S, Baumbach JI, Mineshita M, Miyazawa T. Exhaled breath
26 analysis for lung cancer detection using ion mobility spectrometry. *PLoS One.* 2014;9:
27 e114555.
28
- 29 157. Barash O, Zhang W, Halpern JM, Hua Q-L, Pan Y-Y, Kayal H, et al. Differentiation
30 between genetic mutations of breast cancer by breath volatolomics. *Oncotarget.* 2015;6:
31 44864–44876.
32
- 33 158. Wahl HG, Hoffmann A, Häring H-U, Liebich HM. Identification of plasticizers in medical
34 products by a combined direct thermodesorption–cooled injection system and gas
35 chromatography–mass spectrometry. *J Chromatogr A.* 1999;847: 1–7.
36
- 37 159. de Gennaro G, Dragonieri S, Longobardi F, Musti M, Stallone G, Trizio L, et al. Chemical
38 characterization of exhaled breath to differentiate between patients with malignant
39 pleural mesothelioma from subjects with similar professional asbestos exposure. *Anal*
40 *Bioanal Chem.* 2010;398: 3043–3050.
41
- 42 160. Liu H, Wang H, Li C, Wang L, Pan Z, Wang L. Investigation of volatile organic
43 metabolites in lung cancer pleural effusions by solid-phase microextraction and gas
44 chromatography/mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.*
45 2014;945-946: 53–59.
46
- 47 161. Chen X, Xu F, Wang Y, Pan Y, Lu D, Wang P, et al. A study of the volatile organic
48 compounds exhaled by lung cancer cells in vitro for breath diagnosis. *Cancer.* 2007;110:
49 835–844.
50
- 51 162. Leopold JH, Bos LDJ, Sterk PJ, Schultz MJ, Fens N, Horvath I, et al. Comparison of
52 classification methods in breath analysis by electronic nose. *J Breath Res.* 2015;9:
53 046002.
54
- 55 163. Khoubnasabjafari M, Mogaddam MRA, Rahimpour E, Soleymani J, Saei AA, Jouyban A.
56 *Breathomics: Review of Sample Collection and Analysis, Data Modeling and Clinical*
57 *Applications.* *Crit Rev Anal Chem.* 2021; 1–27.
58
59
60

- 1
2
3 164. Politi L, Monasta L, Rigressi MN, Princivale A, Gonfiotti A, Camiciottoli G, et al.
4 Discriminant Profiles of Volatile Compounds in the Alveolar Air of Patients with
5 Squamous Cell Lung Cancer, Lung Adenocarcinoma or Colon Cancer. *Molecules*. 2021.
6 p. 550. doi:10.3390/molecules26030550
7
- 8 165. Tassopoulos CN, Barnett D, Fraser TR. Breath-acetone and blood-sugar measurements
9 in diabetes. *Lancet*. 1969;1: 1282–1286.
10
- 11 166. Guo D, Zhang D, Zhang L, Lu G. Non-invasive blood glucose monitoring for diabetics by
12 means of breath signal analysis. *Sens Actuators B Chem*. 2012;173: 106–113.
13
- 14 167. Pleil JD, Beauchamp JD, Dweik RA, Risby TH. A special issue: Flow, pressure, volume
15 and time as dependent variables in breath analysis. *J Breath Res*. 2020;15: 010201.
16
- 17 168. Sylvester KP, Clayton N, Cliff I, Hepple M, Kendrick A, Kirkby J, et al. ARTP statement
18 on pulmonary function testing 2020. *BMJ Open Respir Res*. 2020;7.
19 doi:10.1136/bmjresp-2020-000575
20
- 21 169. Pleil JD, Ariel Geer Wallace M, Davis MD, Matty CM. The physics of human breathing:
22 flow, timing, volume, and pressure parameters for normal, on-demand, and ventilator
23 respiration. *J Breath Res*. 2021;15. doi:10.1088/1752-7163/ac2589
24
- 25 170. Ibrahim W, Cordell RL, Wilde MJ, Richardson M, Carr L, Sundari Devi Dasi A, et al.
26 Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography-mass
27 spectrometry. *ERJ Open Res*. 2021;7. doi:10.1183/23120541.00139-2021
28
- 29 171. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline
30 Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to
31 ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323: 1574–1581.
32
- 33 172. Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and
34 Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. *JAMA*.
35 2020;323: 1612–1614.
36
- 37 173. Phillips M, Cataneo RN, Chaturvedi A, Danaher PJ, Devadiga A, Legendre DA, et al.
38 Effect of influenza vaccination on oxidative stress products in breath. *J Breath Res*.
39 2010;4: 026001.
40
- 41 174. Traxler S, Bischoff A-C, Saß R, Trefz P, Gierschner P, Brock B, et al. VOC breath profile
42 in spontaneously breathing awake swine during Influenza A infection. *Sci Rep*. 2018;8:
43 14857.
44
- 45 175. Agarwal SM, Sharma M, Fatima S. VOCC: a database of volatile organic compounds in
46 cancer. *RSC Adv*. 2016;6: 114783–114789.
47
- 48 176. Janfaza S, Banan Nojavani M, Khorsand B, Nikkhah M, Zahiri J. Cancer Odor Database
49 (COD): a critical databank for cancer diagnosis research. *Database* . 2017;2017.
50 doi:10.1093/database/bax055
51
52
53
54
55
56
57
58
59
60

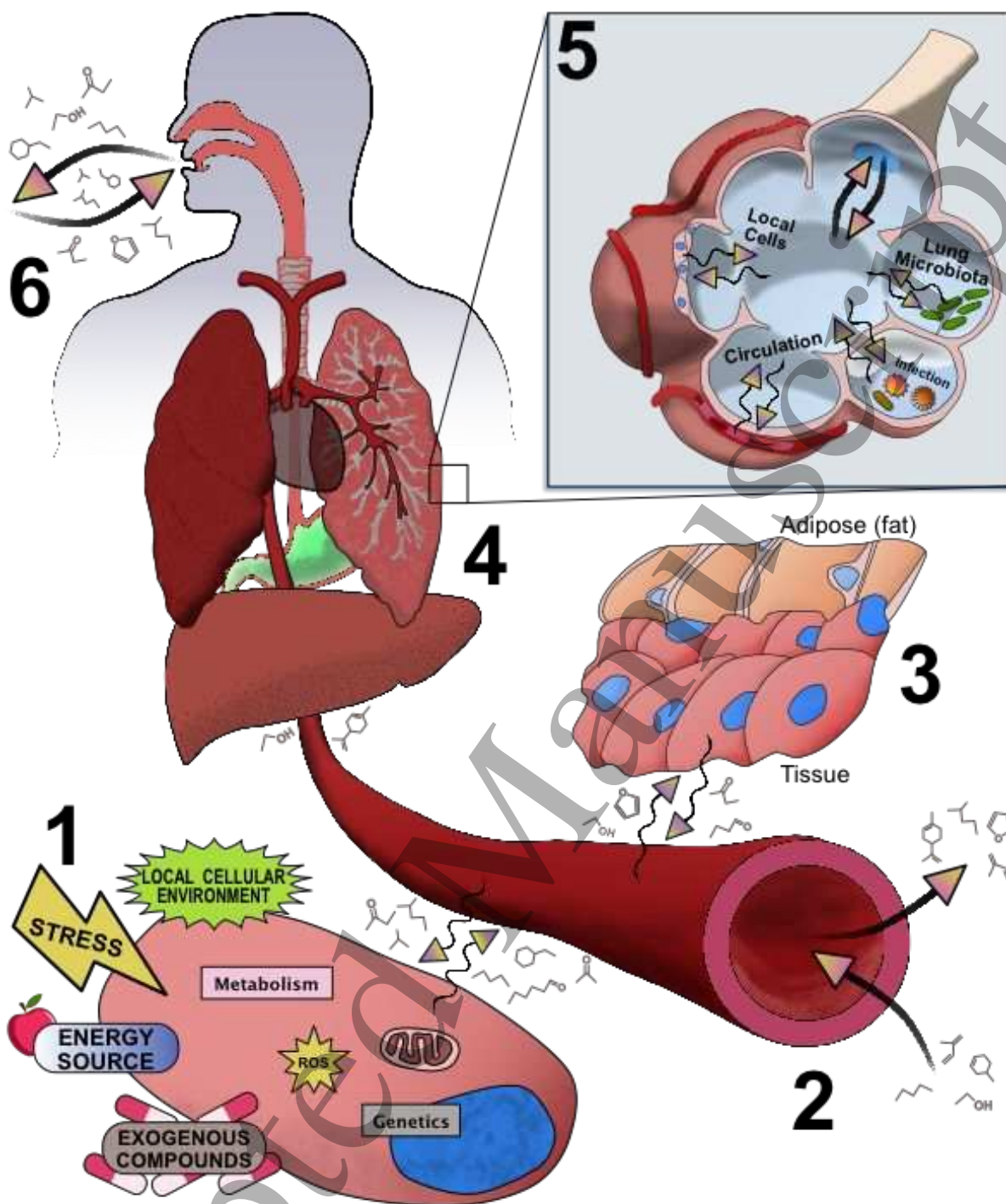


Figure 2. The journey of volatile compounds: from cell to breath. Compounds detected in the breath can be traced back to local cellular changes. Volatiles interact with tissues, organs and other compounds as they move around the body, influencing their expression in the breath. **1)** Local environmental changes and stimulating factors elicit cellular response which in turn alters volatiles both given out and taken up. **2)** Volatile compounds diffuse in and out of the blood stream to move around the body. **3)** As compounds move around the body they diffuse in and out of tissues dependent upon saturation and their affinity for blood, fat or tissue. **4)** Volatile compounds can be metabolised by enzymes such as CYP450s, highly expressed in the liver. **5)** Gases diffuse in and out of the blood in the lung across the alveolar wall. Volatiles from the blood mix with those released by local lung and immune cells, the lung microbiome and infectious bodies. **6)** Compounds are inhaled and exhaled breath is a mixture of alveolar, lung and mouth air with volatiles from the stomach and upper gastrointestinal tract.

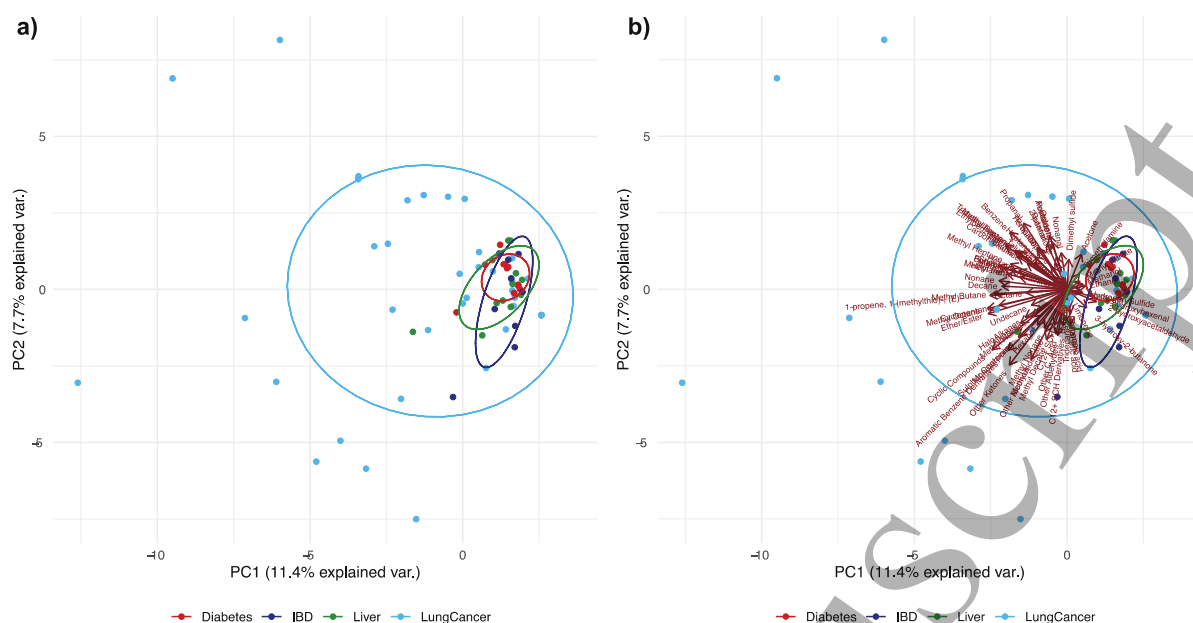


Figure 2. PCA plots of volatiles released by patients for; diabetes (n13), IBD (n9), liver disease (n12) lung cancer (n41). **a**, **b** show same data, axis/compound identifiers shown in **b**. Ellipses represent 95% CI.

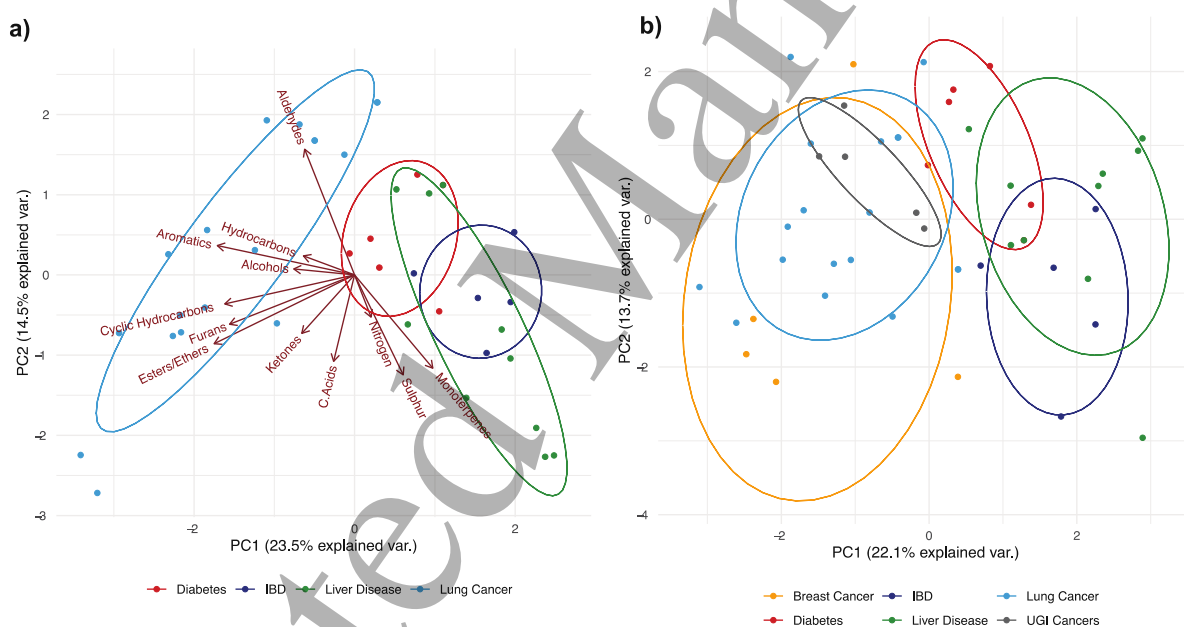


Figure 3. PCA plot of volatiles released by patients, arranged by functional group as shown by axis in **(A)** for; diabetes (n6), IBD (n5), liver disease (n10) and lung cancer (n18). **B**, as **A**, with additional groups; breast cancer (n5) and UGI Cancers (n5). All studies shown are non-exclusionary analytical approaches. Ellipses represent 95% CI.

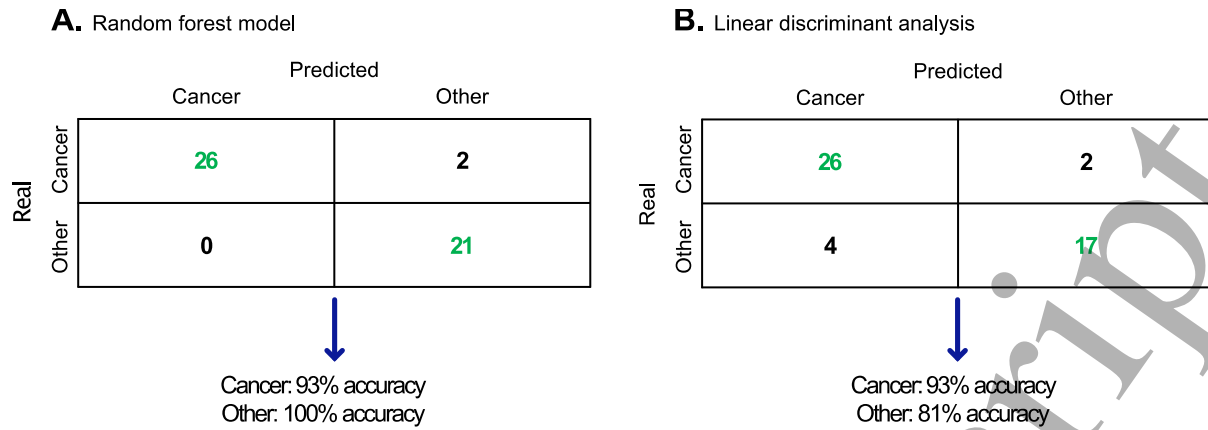


Figure 4. Confusion matrices to summarise supervised classification predictions for **(A)** Random forest model and **(B)** Linear discriminant analysis. Classification accuracy scores for each model are also provided.

Accepted Manuscript