Journal of Breath Research

ACCEPTED MANUSCRIPT

Towards the use of breath for detecting mycobacterial infection: a case study in a Murine model

To cite this article before publication: Flavio Antonio Franchina *et al* 2017 *J. Breath Res.* in press https://doi.org/10.1088/1752-7163/aaa016

TOWARDS THE USE OF BREATH FOR DETECTING MYCOBACTERIAL INFECTION: A CASE STUDY IN A MURINE MODEL

Flavio A. Franchina[1], Theodore R. Mellors[1], Minara Aliyeva[2], Jeff Wagner[3], Nirav Daphtary[2] Lennart K.A. Lundblad[2], Sarah M. Fortune[3], Eric J. Rubin[3], and Jane E. Hill[1,*]

[1]*School of Engineering at Dartmouth College, 14 Engineering Drive, 03755 Hanover (NH), United States*

[2]*University of Vermont, College of Medicine, 149 Beaumont Ave., 05405 Burlington (VT), United States*

[3]*Harvard School of Public Health, Department of Immunology and Infectious Diseases, 02115 Boston (MA), United States*

Manuscript prepared for submission as an original research article to **Journal of Breath Research**

*Corresponding author. *E-mail address*: jane.e.hill@dartmouth.edu

Abstract

In the present research, the potential of breath analysis by comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC×GC-MS) was investigated for the discrimination between healthy and infected mice. A pilot study employing a total of 16 animals was used to develop a method for breath analysis in a murine model for studying MTBC (*Mycobacterium tuberculosis* complex) using the *M. bovis* BCG (Bacillus Calmette-Guérin). Breath was collected in Tedlar bags and concentrated onto thermal desorption tubes for the subsequent analysis by GC×GC-MS. Immunological test and bacterial cell count in bronchoalveolar lavage fluid and mice lung homogenate confirmed the presence of bacteria in the infected group. From the GC×GC-MS analysis, 23 molecules were found to mainly drive the separation between control and infected mice and their tentative identification is provided. This study shows that the overall used methodology is able to differentiate breath between healthy and infected animals and the information herein can be used to further develop the mouse breath model to study MTBC pathogenesis, evaluate pre-clinical drug regimen efficacy, and to further develop the concept of breath-based diagnostics.

Keywords: *Mycobacterium tuberculosis* complex; breath analysis; comprehensive two-dimensional gas chromatography; mass spectrometry; volatile metabolites;

1. Introduction

Since the beginning of recorded history, breath has been used to gain insight into disease pathogenesis. For example, in the Ancient Greece, Hippocrates recognized the diagnostic usefulness of breath, describing *fetor oris* and *fetor hepaticus* (bad breath due to putrid humors in the stomach and by liver failure, respectively) in his treatise on breath aroma and disease [1].

In the contemporary era, Linus Pauling was the first to demonstrate that human breath was composed of many molecules, and not only the classic gases nitrogen, oxygen, carbon dioxide and water vapor. Pauling reported the presence of 250 molecules in exhaled breath using gas chromatography with a flame ionization detector (GC-FID) [2].

Generally, a breath sample is easy to obtain (non-invasive) and contains thousands of molecules, some of which have the potential for use in diagnostic applications [3,4]. To-date, breath analysis has been used to track or diagnose a variety of diseases in humans; for example, in metabolic disorders, monitoring specific metabolites (short chain fatty acids) during a glucose challenge in diabetic patients [5], in lung-related diseases, such as cancer [6], chronic obstructive pulmonary disease [7], and bacterial pneumonias infections (*e.g.* caused by *Staphylococcus aureus*, *Escherichia coli*, and *Candida* spp) [8].

Infectious diseases are the major cause of death in many countries, and earlier diagnosis may result in the earlier initiation of appropriate antimicrobial therapy, as well as limit disease transmission, thus decreasing both morbidity and mortality.

The traditional microbiological diagnostic tests are time-consuming, such as for the common bacterial pathogen *S. aureus*, the gold standard culture approach can take 2-4 days, and even longer (up to 3-6 weeks) for slow-growing bacteria (*i.e.*, *Mycobacterium tuberculosis* complex (MTBC) organisms) [9].

These organisms belonging to MTBC are a group of genetically-related mycobacteria which can cause tuberculosis, which is responsible for considerable morbidity and mortality worldwide, with 10.5 million new cases and 1.5 million deaths in 2015 [10].

The MTBC comprises *M. tuberculosis, M. bovis, M. africanum, M. caprae, M. microti, M. canettii, M. pinnipedii, M. orygis, and M. mungi* and the predominant infection location for these organisms is in the lungs, even though infections at other bodily sites can occur [11].

Breath molecules have been proposed as a way to diagnose MTBC infections and could also be used to track patient treatment response. Seven published studies focusing on *M. tuberculosis* lung infection in humans have analyzed breath collected from over 300 confirmed tuberculosis-positive individuals, and over 1000 patients overall [12–18]. Of note, from these seven studies, only one molecule (tridecane) is shared between two studies and reported as a tuberculosis-associated biomarker. This lack of biomarker consistency is likely due to a series of factors, including the complexity, co-morbidity and metabolic heterogeneity within the human population, as well as the use of different analytical approaches and differences in study design.

In animals, limited studies have been conducted on breath from individuals infected with MTBC. Two studies on bovine breath involved a total of 25 infected animals [19,20] and one non-human primate study involving 5 infected animals [21], showed that healthy and diseased animals could be distinguished using volatile molecules detectable in breath.

Few studies have attempted to link the breath volatile metabolites to disease pathogenesis. To-date, only one study has sought to link mycobacterial metabolism and *in vivo* pathogenesis via human breath [12]. As it is complicated to look for causative links in a heterogeneous study population, such as humans, animal models are generally employed as a first step.

Among these, the murine model is the most commonly used animal model for MTBC infections, due to ease of manipulation and housing, availability of inbred strains or mutants, cost, and genetically-altered strains. To that end, the mouse has been used extensively to model various aspects of MTBC infections, employing mutant strains, immune response, assessing drug efficacy, and new vaccines evaluation [22].

In the present research, a pilot study employing 16 mice was used to develop a method for breath analysis in a murine model for studying MTBC using *M. bovis* BCG (Bacillus Calmette-Guérin), a Biosafety Level 2 organism. Breath was collected in Tedlar bags from anesthetized, intubated mice and concentrated onto thermal desorption tubes for the subsequent analysis by comprehensive 2D gas chromatography coupled to mass spectrometry (GC×GC-MS) [23]. The presence of infection was confirmed with immunological test and bacterial cell count in bronchoalveolar lavage fluid and mice lung homogenate, respectively. A tentative identification of the molecules in breath (sampled at two distinct sampling points) driving the separation between control and infected mice is provided. The information obtained in this study can be used to further develop the mouse breath model for MTBC lung infections.

2.0 Materials and Methods

2.1 Bacterial strain, growth and inoculant preparation

The *M. bovis* BCG substrain was obtained from the Harvard T.H. Chan School of Public Health (Boston, MA), and it was cultured aerobically (21 days; 37°C under agitation) in 20 mL Middlebrook 7H9 (BD Diagnostics, Franklin Lakes, NJ, USA), supplemented with ADC enrichment (BD Diagnostics), 0.05% Tween 20 and 0.2% glycerol, before bacteria were inoculated into the mice airways. For inoculation in murine lung, the cultures were washed 3

times and resuspended in phosphate-buffered saline (PBS, pH = 7.4) to the desired final concentration.

2.2 Mice and infection route

All mice were housed in the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-accredited animal facility at the University of Vermont (Burlington, VT). The protocol for animal infection and respiratory physiology measurements was approved by the Institutional Animal Care and Use Committee, in accordance with AAALAC guidelines. Male C57BL/6J mice (10 – 12 week-old) were purchased from The Jackson Laboratories (Bar Harbor, ME). Mice were exposed to 50 μL inoculum in the 10^5-10^6 CFU (colony-forming unit) range of *M. bovis* BCG cell via oropharyngeal aspiration, as previously reported [24], and breath was collected at either 24h (n = 4 per group) and 48h (n = 4 per group) post-inoculation. The control group were inoculated with 50 μL of PBS and treated identically. A total of 16 mice were used.

2.3 Exhaled breath collection

Twenty-four and forty-eight hours after the *M. bovis* BCG or PBS instillation, mice were anesthetized with pentobarbital and their tracheas cannulated. The mice were placed on the ventilator (flexiVent, SCIREQ, Montreal, QC, Canada). Breath from the ventilator was collected in 1L Tedlar bags (SKC, Eighty Four, PA) at 200 breaths/min with a positive end expiratory pressure (PEEP) of 3 cm H_2O for 30 – 40 min [24]. Room air samples (n = 16) were collected by attaching the Tedlar bag to the ventilator in absence of mouse. Only new Tedlar bags after the

conditioning step were used for collecting breath and room air. The conditioning step consisted of flushing the bags with UHP nitrogen gas and condition them in an incubator at 37°C for 30 minutes. This process was repeated 3 times.After bag filling, the breath or room air volatile molecules were concentrated by drawing the content of the Tedlar bag onto thermal desorption tubes packed with Carbopack Y, Carbopack X and Carboxen 1000 (Supelco, Bellefonte, PA) via a vacuum pump, at an average flow rate of 170 mL/min. The sampling flow rate was measured during each tube loading by means of a flowmeter and the total drawing time was adjusted to have a constant final volume of *circa* 1 L.

The loading onto thermal desorption tubes was completed within five min of breath collection. Tubes were stored at 4°C and analyzed in the GC×GC-MS within two weeks of collection.

2.4 Collection and analysis of bronchoalveolar lavage fluid and lung tissue

After the breath collection, 1 mL of room temperate PBS was instilled into the lungs via the tracheal cannula and then withdrawn. The BALF (bronchoalveolar lavage fluid) was centrifuged and the supernatant collected. The remaining cell pellet was resuspended in 400 µl of PBS and an aliquot (150 µl) of the bronchoalveolar lavage was run on the Guava easyCyte 6HT-2L benchtop flow cytometer. Cell counts were calculated using two parameters, the (A) concentration (particles/ml) and (B) cell percentage gated according to particle size (forward scatter, FSC) and granularity (side scatter, SSC) to exclude RBCs (red blood cells) and debris. The following calculation was applied: A*(B*0.01)*0.4= absolute total cell count. The cell concentration of the BALF cells was adjusted to 200 μl^{-1} and 100 µl was loaded into Cytospin funnels and spun onto glass slides at 600 RPM for 6 minutes. Following drying the cells were stained using a HEMA 3 kit (Fisher Scientific); first fixated in a methanol solution for 2 minutes then stained with Eosin

Y for 2 minutes, and finally in Azure A and methylene blue for 30 seconds. The cells were differentiated by counting 300 cells/slide differentiating into eosinophils, neutrophils, macrophages, and lymphocytes.

After the bronchoalveolar lavage was obtained lungs were harvested and homogenized. The excised lungs were trimmed free from adjacent tissue and the lung lobes placed in sterile gentleMACS M tubes (Miltenyi Biotec, Auburn, CA) with 3 ml of PBS. The lungs were homogenized by running the RNA_01 program on a gentleMACS dissociator. Lung bacterial cell counts were obtained after 21 days at 37°C by plating an aliquot of the lung homogenate on 7H10 Middlebrook agar plates supplemented with OADC (BD Diagnostics).

The bronchoalveolar lavage fluid was centrifuged, the resulting pellet was cultured neat (no dilutions), and the presence of visible colonies was assessed after three weeks as an additional confirmatory test for the presence of bacterial cells in the lungs at the moment of breath collection. Bronchoalveolar lavage fluid cell differentiation and cell count were performed in triplicate.

2.5 Analytical instrumentation and statistical analysis

Thermal desorption tubes were analyzed in a Pegasus 4D (LECO Corporation, St. Joseph, MI) GC×GC time-of-flight (TOF) MS instrument with an Agilent 7890 GC equipped with a TDU, CIS and an MPS auto sampler (Gerstel, Linthicum Heights, MD). Details of the sample desorption, chromatography and mass spectrometry experimental conditions are summarized in **Table 1**. Data acquisition and analysis were performed using ChromaTOF software, version 4.50 (LECO Corp.).

Chromatographic data were processed and aligned using ChromaTOF. For peak identification, a signal-to-noise (S/N) cutoff was set at 20, and resulting peaks were identified by a forward search of the NIST 2011 library. For tentative peak identification, a forward match score of 800 (of 1000) was required. For the alignment of peaks across chromatograms, maximum first and second-dimension retention time deviations were set at 6s and 0.1s, respectively, and the inter-chromatogram spectral match threshold was set at 700.

All statistical analyses were performed using R (version 3.3.0). Prior to statistical analyses, samples were normalized using probabilistic quotient normalization [25]. After the manual removal of artifacts (see supplementary Table S1), a further feature reduction step was performed, removing features present in less than 50% of the chromatograms in the *M. bovis* BCG infected, PBS control mice or room air groups. Random forest was then applied for the selection of features that discriminate between these three groups [26]. Features were selected as discriminatory if they were ranked in the top 100 in 100/100 random forest iterations of 1000 trees, based on their mean decrease in impurity. Principal component analysis (PCA) was used to visualize the variance between samples when looking at the selected discriminatory features [26]. Centroid analysis was used to visualize the center of the *M. bovis* BCG, PBS control, and room air clusters based on their principal component (PC) scores [26]. The significance of the compounds was tested by means of the Mann-Whitney U test.

Table 1. Summary of GC×GC-TOFMS experimental conditions.

Thermal desorption tubes and desorption process	
Sorbent material	Carbopack Y, Carbopack X, Carboxen 1000
Solvent venting time (TDU temperature; flow rate)	10 min (30°C; 60 ml/min)
Thermal desorption time (TDU temperature; flow rate)	5 min (330°C; 60 ml/min)
Cryofocusing time (CIS temperature)	5 min (-120°C)
Sample desorption time, CIS temperature, injection mode	180 s, 330 °C, splitless
GC×GC	
D1 column (L × ID × d_f)	Rxi-624Sil (60 m × 0.25 mm × 1.4 µm)
D2 column (L × ID × d_f)	Stabilwax (1 m × 0.25 mm × 0.5 µm)
D1 oven temperature	35 to 235°C at 3.5 °C/min
D2 oven and modulator temperature	+ 5°C and + 25 °C relative to D1 oven temp
Modulation period (hot/cold jet)	2.5 s total (alternating 0.85 s/0.4 s hot/cold)
Carrier gas (flow rate)	Helium (2 mL/min)
Transfer line temperature	260 °C
MS	
Acquisition mass range	30 - 400 *m/z*
Acquisition frequency	200 Hz
Ion source temperature	200 °C

3.0 Results and Discussion

3.1 Murine infection confirmation

To verify the presence of bacteria at the time of breath sampling, cell count was obtained after plating homogenized lungs and bronchoalveolar lavage fluid. For mice infected with *M. bovis* BCG, bacterial counts from homogenized lungs at the time of breath sampling ranged between 1 × 10^4 and 1 × 10^6 CFU/mouse and were not statistically different (p = 0.35) between 24 h and 48 h sampling points. Mice inoculated with the phosphate-buffer saline did not yield any bacterial colonies from either lung homogenate or bronchoalveolar lavage fluid (**Supplementary Figure S1**).

Additionally, bronchoalveolar lavage fluid from mice inoculated with *M. bovis* BCG cultured positive while mice inoculated with PBS cultured negative, confirming the homogenate results. To further confirm infection status, host immune cell differentiation in the bronchoalveolar lavage fluid was measured. No eosinophil and lymphocytes were found in any of the mice since they are commonly the predominant cells observed during allergic states [27] and thus no to be expected during bacterial infection. An average (± SD) of 2.2 ± 0.8 × 10^5 neutrophils were

measured in bronchoalveolar lavage fluid of animals exposed to *M. bovis* BCG, which was significantly greater ($p = 0.00094$) than the PBS control group, where none were detected (see **Figure 1**). This is in agreement with other murine studies which highlighted neutrophil influx in the lungs during early active bacterial infections [28,29], including with *M. bovis* BCG after intranasal infection [30]. No significant differences in neutrophil number ($p = 0.88$) were observed between the two harvesting time points.

The average concentration of macrophages in the bronchoalveolar lavage fluid were $1.3 \pm 0.9 \times 10^5$ and $5 \pm 1.4 \times 10^4$ cells for the mice inoculated with BCG and PBS, respectively, and was significantly different ($p = 0.04$) between the two groups (**Figure 1**). No significant difference in macrophages cell number ($p = 0.8$) between the two harvesting time points was observed. The macrophage concentration, both in the infected and control groups, is in agreement with the literature [29,31,32]. Taken together, these biological assays indicate successful infection for the mice inoculated with *M. bovis* BCG.

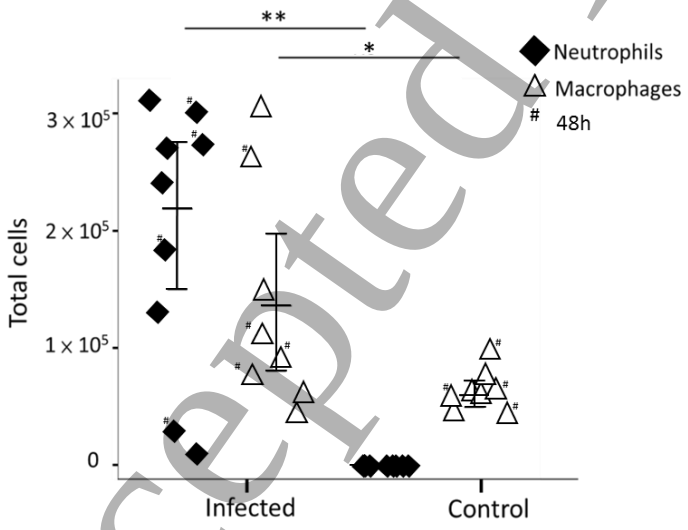


Figure 1. Concentration of neutrophils (solid diamonds) and macrophages (open triangles) in bronchoalveolar lavage fluid collected from infected and control mice ($n = 8$ per group) at either 24 h or 48 h post-inoculation. No eosinophil and lymphocytes are plotted since they were not

present. Error bars represent the 95% interval confidence and significance values were calculated applying the Mann-Whitney significance U-test (*$p < 0.05$; **$p < 0.01$).

3.2. Breathprint of control and infected mice

Breath from mice inoculated with *M. bovis* BCG and PBS ($n = 8$ per group), as well as room air samples ($n = 16$), were analyzed using GC×GC-TOFMS. The flowchart of the data reduction steps adopted in the current study is shown in **Supplementary Figure S2**.

Across the 32 samples, approximately 1600 peaks were detected. In a preliminary data cleaning step, 300 peaks were removed based on mass spectrum [33] and presence in all samples (both room air and breath were collected in Tedlar bags). A spectral similarity to known environmental contaminants or artifacts coming from the Tedlar bags (*e.g.*, phenol and N,N-dimethylacetamide) or column bleed (*e.g.*, siloxane or silyl derivatives), in addition to their persistent presence in all the chromatograms, drove the exclusion criteria (for a more detailed list of excluded features see Supplementary Table S1).

In order to direct data analysis to consistently detected peaks, those present in more than 50% of samples from each group (infected, control, and room air) were considered for further analysis, resulting in 604 features.

Bubble plots representing the two-dimensional GC separation obtained from the infected and control mouse breath, and the room air samples are reported in **Figure 2a-c**, respectively. Each bubble represents a chromatographic peak and bubble position is given by the retention time coordinates of the peak (*i.e.*, first and second dimension retention time). The size of the bubble corresponds to the peak area calculated from the deconvoluted MS signal.

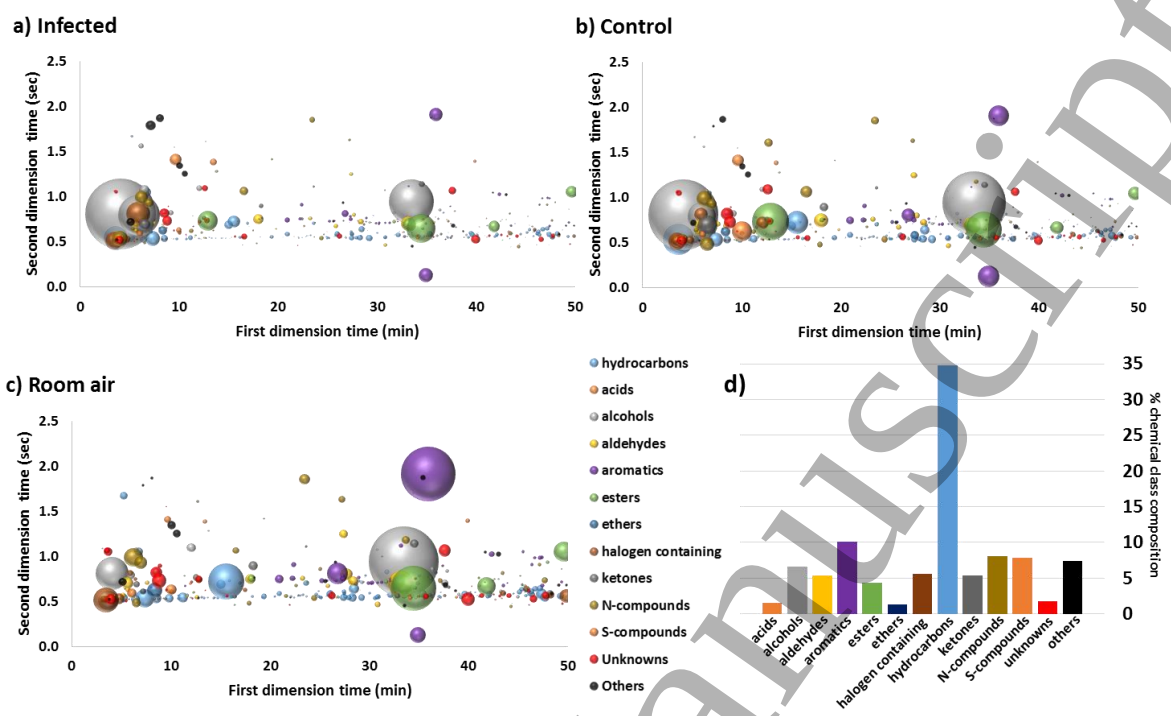


Figure 2a-d. Bubble plot of the GC×GC chromatograms of breath from a) mice infected with *M. bovis* BCG, b) mice inoculated with PBS, and c) room air samples. Each bubble represents a detected peak and the size is proportional to the average area across chromatograms belonging to that sample group. Only features (604) present in 50% of the samples from single class are plotted. d) Chemical class relative abundance based on number of identified compounds, not on their relative concentration.

The profile of the 604 features was found to have a moderate correlation among the samples analyzed, with a Spearman's rank correlation of 0.53, 0.5, and 0.52 for the mice treated with *M.bovis* BCG, PBS, and room air samples, respectively [34].

Most of the 604 peaks were assigned chemical classification based on their mass spectral similarities (≥800) to the NIST 2011 library (**Figure 2a-d**). A wide range of chemical classes was found (*i.e.*, carboxylic acids, alcohols, aldehydes, aromatic, halogen-containing, N-containing, S-containing compounds, ethers, esters, ketones) with the hydrocarbons being the most commonly present chemical class (34%) (**Figure 2d**). The other remaining chemical

classes ranged from 2-10 % of breath composition. These classes have been previously reported in the breath of animals and humans, with the hydrocarbons being the most abundant [4,12–21].

3.3. Selected molecules in breath for infected and control mice discrimination

From the 604 detected breath molecules, 23 features were selected by random forest that discriminated between mice infected with *M. bovis* BCG, mice inoculated with PBS, and room air samples. When considering these 23 selected features, the correlation between breath samples of the same class is highly correlated, with a Spearman's rank correlation of 0.7, 0.57, and 0.55 for the mice treated with *M. bovis* BCG, PBS, and room air samples, respectively [34]. The lower correlation between room air samples when considering these 23 selected features is possibly due to variations in room air composition on different days of sampling.

Principal components scores were used to visualize variance between these sample types (**Figure 3**). The first two components explain 53% of the variance between the three group types. Data centroids (**Figure 3**, circles) illustrate the average position on component 1 and component 2 of all the samples for each group. The Euclidean distance between breath (infected and control mice) and room air was found to be 1.9 times greater than the distance between BCG and PBS clusters. Although differences in profile are observed between these three groups when considering the 23 selected molecules, the small sample size of the present study prevented external validation of these preliminary findings. Therefore, future studies are necessary to evaluate the reproducibility of these identified biomarkers. Both the 24 h and 48 h breath samples cluster within their respective groups, suggesting that the discriminatory features of the breath profile are maintained across the two sampling points. One room air sample is misclassified, falling within the control group sample projections. All the room air samples

collected from the same day cluster with their respective and expected group, therefore, the reason for this misclassification is unclear.

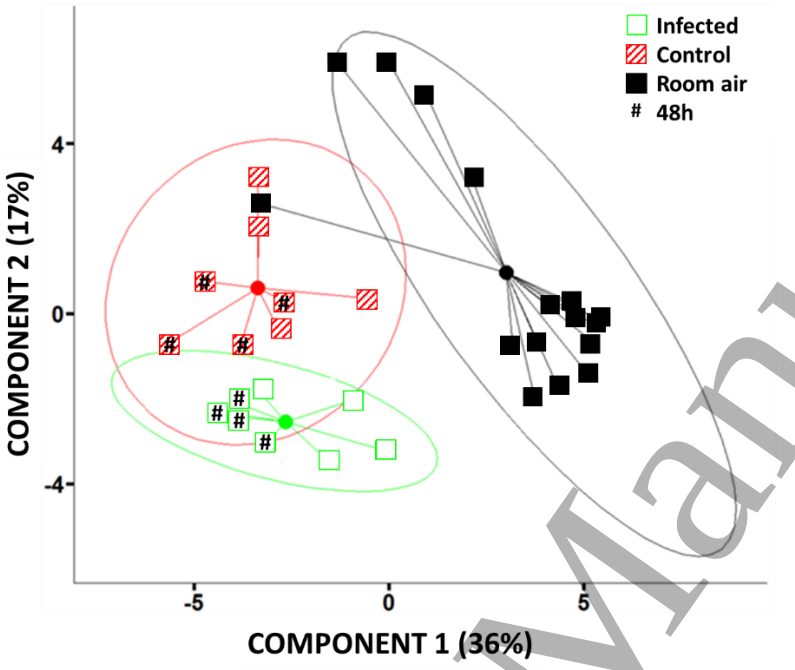


Figure 3. Centroid analysis using principal component scores of the *M. bovis* BCG infected mice (red striped square), PBS-inoculated mice (green empty square) and the room air (black solid square), considering the 23 discriminatory features. Ellipses represent the 95% confidence interval for each class.

Tentative identifications were provided for all 23 features based only on mass spectral matches to the NIST 2011 library (≥800) and are reported in **Table 2**.

These molecules belong to eight different chemical classes (nine hydrocarbons, four aldehydes, one N-containing and two S-containing compounds, two alcohols, one ester, two aromatics, and two ketones). Two of these molecules (2-methyl-(E)-2-butenal and 2-ethyl-3-methylbutanal) were exclusively detected in breath (*i.e.*, not detected in any room air samples) while 21 were present in all samples but at different concentrations. It should be pointed out that even if MS

spectra library comparison commonly used in GC-MS applications for peak identification, the final confirmation should assessed by injecting analytical standards.

Based on principal components loading scores, the molecules that mainly drive the separation between infected and control mice were: 4-methyldecane, 2,4,4-trimethyl-1-pentene, butanal, 3-oxo cyclopenten-1-yl-(2E)-penta-2,4-dienoate, acetaldehyde, decyl-cyclohexane, acetone and 3,4-diethyl-2-hexane (**Supplementary Figure S3**). Only decyl-cyclohexane was significantly more abundant in infected mice compared with the control mice treated with PBS ($p < 0.05$). These molecules were detected in breath of both infected and control mice, but they were differentially abundant (**Table 2**). A connection between these molecules and specific metabolic pathways has yet to be determined.

To the best of the authors' knowledge, this study represents the first time breath has been evaluated from mice infected with any MTBC organisms. In considering previous studies of exhaled breath in the context of mycobacteria infection, there are few, and they are limited to cows [19,20], non-human primates [21], and humans [12–18]. Beyond the use of different models, there are some important differences in the studies, including the I) *Mycobacterium* species and strain used (*M. bovis* in cows, *M. tuberculosis* Erdman in non-human primates, unknowns *M. tuberculosis* strain(s) in humans, vs. *M. bovis* BCG in the present study), II) longevity of infection (3 months (cows), 1-2 months (non-human primates), unknown (humans), vs. 24-48 h in the present study), and III) breath volume collected (1.5L (cows and non-human primates), 1L (humans), vs. 1L in the present study). In addition, each study used different sorbent phases to concentrate and store breath. One study [21] used the same analytical technique as this study, however, the others all employed one-dimensional GC-MS systems.

Considering all these different factors, a comparison of the molecules from this study with others is challenging.

However, in an effort to translate this work to others, a remark on the overlap between the molecules detected in the present and those reported in previous studies is provided. Eight molecules detected in mouse breath were previously reported in the breath of infected non-human primates and cows. Four molecules (tridecane, 2-heptanone, acetic acid phenyl ester and *o*-cymene) were detected in the breath of non-human primates and reported as discriminatory and statistically different between healthy and infected [21]; and four other molecules (styrene, benzaldehyde, acetophenone, and nonanal) were reported in the breath of cows [19,20]. In the present study, these molecules were not a part of the discriminatory feature set and they were not statistically different between infected and control mice.

Human breath studies [12,14,18] cumulatively reported on 48 molecules, that discriminated between patients whose sputum cultures were positive or negative for mycobacteria.

In the present study, four of the 48 (3-methylheptane, ethylbenzene, 4-methyldecane, and acetaldehyde) were detected in mice breath and result discriminatory (**Table 2**). Acetaldehyde was detected at a higher abundance in the mice infected with *M. bovis* BCG while 3-methylheptane, ethylbenzene, and 4-methyldecane were in higher abundance in the PBS group.

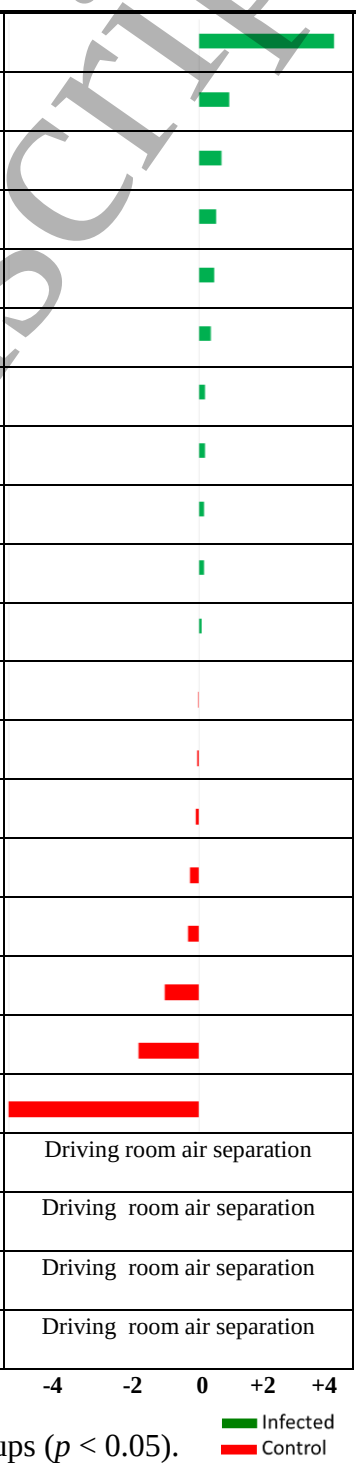
Nine other molecules from the aforementioned human studies (4,6,8-trimethyl-1-nonene, tridecane, octanal, 1,3-dimethyl-cyclohexane, camphene, heptanal, nonanal, 2-methylpentane, styrene) were also detected in the breath of mice in the present contribution, however, none of these molecules were discriminatory or statistically different between groups.

A comparison with volatile molecules from mycobacterial culture [35,36] is even more complicated, due to differences in the sampling technique (*i.e.* SPME or NTME vs TD), and

heterogeneity of the biological conditions (*i.e.* different species of mycobacteria, *in vitro* vs *in vivo* environment *etc*.).

Table 2. Tentative peak identification of the discriminatory molecules selected by random forest, formula, and relative abundance between the control and infected mice groups.

Peak no.	Tentative compound ID	Formula	Log_2(Control/Infected)
1	decyl-cyclohexane*	$C_{16}H_{32}$	
2	methanethiol	CH_4S	
3	3,4-diethyl-2-hexene	$C_{10}H_{20}$	
4	dimethyl sulphite	$C_2H_6O_3S$	
5	2,2,4,4-tetramethyl-3-pentanone	$C_9H_{18}O$	
6	acetaldehyde	C_2H_4O	
7	acetone	C_3H_6O	
8	2-butene	C_4H_8	
9	3-oxo cyclopenten-1-yl-(2E)-penta-2,4-dienoate	$C_{10}H_{10}O_3$	
10	4-methyl-decane	$C_{11}H_{24}$	
11	ethylbenzene	C_8H_{10}	
12	2,4,4-trimethyl-1-pentene	C_8H_{16}	
13	4-methylene-1-(1-methylethyl)-bicyclo[3.1.0]hexane	$C_{10}H_{16}$	
14	2-methyl-(E)-2-butenal	C_5H_8O	
15	2-ethyl-3-methylbutanal	$C_7H_{14}O$	
16	2-methylheptane	C_8H_{18}	
17	butanal	C_4H_8O	
18	3-methylheptane	C_8H_{18}	
19	(2-aziridinylethyl)amine	$C_4H_{10}N_2$	
20	3-methyl-(Z)-4-nonene	$C_{10}H_{20}$	Driving room air separation
21	2-decen-1-ol	$C_{10}H_{20}O$	Driving room air separation
22	4-[2-(methylamino)ethyl]-phenol	$C_{10}H_{15}NO_2$	Driving room air separation
23	3,4-dimethylpentanol	$C_7H_{16}O$	Driving room air separation



*Feature statistically different between infected and control mouse groups ($p < 0.05$).

Conclusions and future perspectives

Breath holds great potential as a carrier of information on the health status of an individual and animal models are important tools in the quest to better diagnose and understand infectious diseases.

In this study, breath was collected for the first time from mice infected with an MTBC pathogen, *M. bovis* BCG, and contrasted with breath from animals instilled with PBS. The immunological and bacterial cell count data at the time of breath collection confirmed the infection status of the animals treated with the pathogen. From the entire spectrum of volatile metabolites in breath, twenty-three molecules were tentatively identified and selected as a panel that could distinguish between infected and uninfected animals. Four of them are in common with the possible biomarkers reported in previous studies on breath from animals with an MTBC infection.

The information obtained in this study can be used to further develop a mouse breath model for MTBC lung infections as potential diagnostic or drugs treatment monitoring approach, as well as a tool for a better understanding of disease pathogenesis by volatile metabolites expression. Future studies will focus on establishing links between the metabolites and their associated pathogen and/or host pathways.

Acknowledgements

The authors would like to thank Emily M. Nakada, Ph.D. (University of Vermont) for assisting with flow cytometry. We thank Mavra Nasir and Chris Rees for their aid in editing this manuscript. This work was supported by the Bill and Melinda Gates Foundation (OPP1135516).

References

[1] Adams F 1994 Hippocratic Writings *Hippocratic Writings: Aphorism 4,5* (New York: Web Atomics) p Aphorism 4,5

[2] Pauling L, Robinson A B, Teranishi R and Cary P 1971 Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography *Proc. Natl. Acad. Sci.* **68** 2374–6

[3] Boots A W, Joep J B N van B, Jan W D, Agnieszka S, Emile F W and Frederik J van S 2012 The versatile use of exhaled volatile organic compounds in human health and disease *J. Breath Res.* **6** 27108

[4] de Lacy Costello B, Amann a, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D and Ratcliffe N M 2014 A review of the volatiles from the healthy human body. *J. Breath Res.* **8** 14001

[5] Gruber B, Keller S, Groeger T, Matuschek G, Szymczak W and Zimmermann R 2016 Breath gas monitoring during a glucose challenge by a combined PTR-QMS/GC×GC-TOFMS approach for the verification of potential volatile biomarkers *J. Breath Res.* **10** 36003

[6] Capuano R, Santonico M, Pennazza G, Ghezzi S, Martinelli E, Roscioni C, Lucantoni G, Galluccio G, Paolesse R, Di Natale C and D'Amico A 2015 The lung cancer breath signature: a comparative analysis of exhaled breath and air sampled from inside the lungs *Sci. Rep.* **5** 16491

[7] Fens N, Zwinderman A H, Van Der Schee M P, De Nijs S B, Dijkers E, Roldaan A C, Cheung D, Bel E H and Sterk P J 2009 Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma *Am. J. Respir. Crit. Care Med.* **180**

1076–82

[8] Filipiak W, Beer R, Sponring A, Filipiak A, Ager C, Schiefecker A, Lanthaler S, Helbok R, Nagl M, Troppmair J and Amann A 2015 Breath analysis for *in vivo* detection of pathogens related to ventilator-associated pneumonia in intensive care patients: a prospective pilot study *J. Breath Res.* **9** 16004

[9] Connie R. Mahon, Donald C. Lehman G M J 2008 *Textbook of Diagnostic Microbiology* ed E H Sciences

[10] WHO 2016 *WHO | Global tuberculosis report 2016*

[11] Rodriguez-Campos S, Smith N H, Boniotti M B and Aranaz A 2014 Overview and phylogeny of Mycobacterium tuberculosis complex organisms: Implications for diagnostics and legislation of bovine tuberculosis *Res. Vet. Sci.*

[12] Phillips M, Cataneo R N, Condos R, Ring Erickson G A, Greenberg J, La Bombardi V, Munawar M I and Tietje O 2007 Volatile biomarkers of pulmonary tuberculosis in the breath *Tuberculosis* **87** 44–52

[13] Phillips M, Basa-Dalay V, Blais J, Bothamley G, Chaturvedi A, Modi K D, Pandya M, Natividad M P R, Patel U, Ramraje N N, Schmitt P and Udwadia Z F 2012 Point-of-care breath test for biomarkers of active pulmonary tuberculosis *Tuberculosis* **92** 314–20

[14] Phillips M, Basa-Dalay V, Bothamley G, Cataneo R N, Lam P K, Natividad M P R, Schmitt P and Wai J 2010 Breath biomarkers of active pulmonary tuberculosis *Tuberculosis* **90** 145–51

[15] Sahota A S, Gowda R, Arasaradnam R P, Daulton E, Savage R S, Skinner J R, Adams E, Ward S A and Covington J A 2016 A simple breath test for tuberculosis using ion mobility: A pilot study *Tuberculosis* **99** 143–6

[16] Soobratty M R, Whitfield R, Subramaniam K, Grove G, Carver A, O'Donovan G V., Wu H H T, Lee O Y C, Swaminathan R, Cope G F and Milburn H J 2014 Detecting active pulmonary tuberculosis with a breath test using nanomaterialbased sensors *Eur. Respir. J.* **43** 1519–22

[17] Bruins M, Rahim Z, Bos A, Van De Sande W W J, Endtz H P and Van Belkum A 2013 Diagnosis of active tuberculosis by e-nose analysis of exhaled air *Tuberculosis* **93** 232–8

[18] Kolk A H J, van Berkel J J B N, Claassens M M, Walters E, Kuijper S, Dallinga J W and van Schooten F J 2012 Breath analysis as a potential diagnostic tool for tuberculosis. *Int. J. Tuberc. lung Dis. Off. J. Int. Union against Tuberc. Lung Dis.* **16** 777–82

[19] Peled N, Ionescu R, Nol P, Barash O, McCollum M, Vercauteren K, Koslow M, Stahl R, Rhyan J and Haick H 2012 Detection of volatile organic compounds in cattle naturally infected with Mycobacterium bovis *Sensors Actuators, B Chem.* **171**–**172** 588–94

[20] Ellis C K, Stahl R S, Nol P, Waters W R, Palmer M V., Rhyan J C, VerCauteren K C, McCollum M and Salman M D 2014 A pilot study exploring the use of breath analysis to differentiate healthy cattle from cattle experimentally infected with Mycobacterium bovis *PLoS One* **9**

[21] Mellors T R, Blanchet L, Flynn J L, Tomko J, O'Malley M, Scanga C A, Lin P L and Hill J E 2017 A new method to evaluate macaque health using exhaled breath: A case study of M. tuberculosis in a BSL-3 setting *J. Appl. Physiol.* jap.00888.2016

[22] Flynn J L 2006 Lessons from experimental Mycobacterium tuberculosis infections **8**

[23] Tranchida P Q, Franchina F A, Dugo P and Mondello L 2016 Comprehensive two-dimensional gas chromatography-mass spectrometry: Recent evolution and current trends *Mass Spectrom. Rev.* **35**

[24] Zhu J, Jimenez-Diaz J, Bean H D, Daphtary N A, Aliyeva M I, Lundblad L K A and Hill J E 2013 Robust detection of P. aeruginosa and S. aureus acute lung infections by secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting: from initial infection to clearance. *J. Breath Res.*

[25] Dieterle F, Ross A, Schlotterbeck G and Senn H 2006 Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics *Anal. Chem.*

[26] Hastie T, Tibshirani R and Friedman J 2009 The Elements of Statistical Learning *The Elements of Statistical Learning*

[27] Li S, Aliyeva M, Daphtary N, Martin R a., Poynter M E, Kostin S F, van der Velden J L, Hyman A M, Stevenson C S, Phillips J E and Lundblad L K a. 2014 Antigen-induced mast cell expansion and bronchoconstriction in a mouse model of asthma *AJP Lung Cell. Mol. Physiol.*

[28] Balamayooran G, Batra S, Fessler M B, Happel K I and Jeyaseelan S 2010 Mechanisms of neutrophil accumulation in the lungs against bacteria *Am. J. Respir. Cell Mol. Biol.* **43** 5–16

[29] Bean H D, Jiménez-Díaz J, Zhu J and Hill J E 2015 Breathprints of model murine bacterial lung infections are linked with immune response *Eur. Respir. J.*

[30] Lombard R, Doz E, Carreras F, Epardaud M, Le Vern Y, Buzoni-Gatel D and Winter N 2016 IL-17RA in non-hematopoietic cells controls CXCL-1 and 5 critical to recruit neutrophils to the lung of mycobacteria-infected mice during the adaptive immune response *PLoS One* **11**

[31] Schreiber O, Steinwede K, Ding N, Srivastava M, Maus R, Länger F, Prokein J, Ehlers S,

Welte T, Gunn M D and Maus U a 2008 Mice that overexpress CC chemokine ligand 2 in their lungs show increased protective immunity to infection with Mycobacterium bovis bacille Calmette-Guérin. *J. Infect. Dis.*

[32] Srivastava M, Meinders A, Steinwede K, Maus R, Lucke N, Bühling F, Ehlers S, Welte T and Maus U A 2007 Mediator responses of alveolar macrophages and kinetics of mononuclear phagocyte subset recruitment during acute primary and secondary mycobacterial infections in the lungs of mice *Cell. Microbiol.*

[33] Ghimenti S, Lomonaco T, Bellagambi F G, Tabucchi S, Onor M, Trivella M G, Ceccarini A, Fuoco R and Di Francesco F 2015 Comparison of sampling bags for the analysis of volatile organic compounds in breath *J. Breath Res.*

[34] Hinkle D E, Wiersma W and Jurs S G 2003 *Applied Statistics for the Behavioral Sciences* (Houghton Mifflin)

[35] Küntzel A, Fischer S, Bergmann A, Oertel P, Steffens M, Trefz P, Miekisch W, Schubert J K, Reinhold P and Köhler H 2016 Effects of biological and methodological factors on volatile organic compound patterns during cultural growth of *Mycobacterium avium* ssp . *paratuberculosis J. Breath Res.*

[36] Mellors T R, Rees C A, Wieland-Alter W F, von Reyn C F and Hill J E 2017 The volatile molecule signature of four mycobacteria species *J. Breath Res.* **11** 31002