Volatile Compound Analysis of Breath from Patients with *Clostridium difficile* Infection for Biomarker Discovery and Biological Insight

[1]Teny John, [2]Leen Hasan, [3]Kirk Pappan, [3]Owen Birch, [2]David Grove, [2]Gary Procop, [2]Raed Dweik, [3]Billy Boyle, [3]Max Allsworth, [2]Nabin Shrestha

[1] The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, United States; [2] Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio, United States; [3] Owlstone Medical, 183 Cambridge Science Park, Milton Road, Cambridge CB4 0GJ, UK

1. Background and Objective

- *C. difficile* is one of the leading causes of hospital-acquired infective diarrhea, with current or recent antibiotic treatment being the main risk factor. While diarrhea is the main symptom, certain co-morbidities can make CDIs potentially life-threatening [1].
- Compelling evidence for production of distinctive volatile compounds resulting from CDI indicates that breath sampling may be a viable diagnostic alternative.
- The goal of this study was to identify a wider novel selection of exhaled breath VOCs that can discriminate between patients with and without CDI and to obtain biological insight into CDI associations and risk factors. It was based on a cross-sectional study design with one sample collected per patient after diagnosis from a population of hospitalized patients with diarrhea.

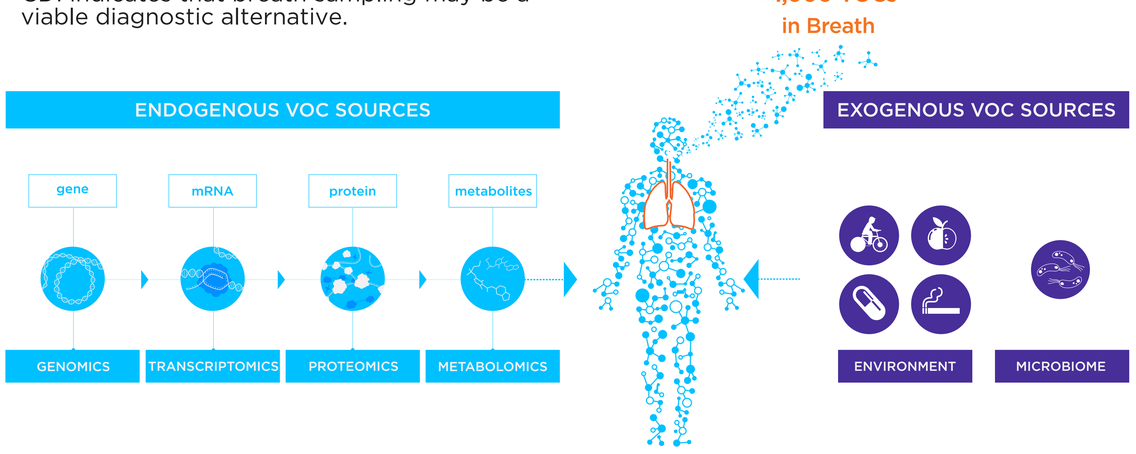

Figure 1: VOCs are low molecular weight metabolites that are excreted in breath as a result of metabolic processes in the body, and can be of endogenous and/or exogenous origin. *Bacterial and viral infections affect the endogenous metabolites and exogenous compounds found in breath.*

2. Methods

- Breath samples were collected at the Cleveland Clinic main hospital from patients breathing into the ReCIVA® Breath Sampler with normal tidal breathing for ~ 10 minutes.
- Breath volatile organic compound (VOC) samples (n=17) were collected from subjects with diarrhea associated with PCR-confirmed *Clostridium difficile* infection (CDI) and matched control subjects with diarrhea but with PCR negative *C. difficile* test results. Breath VOCs are derived from a variety of endogenous and exogenous sources (Fig. 1). A summary of the patient characteristics is given in Table 1.
- Control subjects were recruited with the intention of being the closest demographic match possible to a given CDI case and were tested on the same day as the case study subject.
- Samples were analyzed using the Breath Biopsy Platform in the Breath Biopsy Laboratory (Owlstone Medical Ltd., Cambridge, UK). Untargeted feature extraction was performed for samples that passed all curation checks. Thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS) chromatograms were converted into molecular feature (MF) lists for statistical analysis.
- Principal component analysis (PCA) was used to visually check for any underlying structure in the data. Paired testing was performed using the Wilcoxon sign-rank test and the Mann-Whitney U-test was used for unpaired group comparisons.
- For MFs of interest, a tentative ID was assigned by comparison to the NIST (National Institute of Standards and Technology) library.

Clinical Variable	Controls	Cases	Full Cohort
Number of patients	17	17	34
Male/Female	9/8	8/9	17/17
Age	55.06 ± 13.49	59.71 ± 16.70	57.38 ± 14.91
BMI	29.23 ± 6.91	27.24 ± 6.99	28.24 ± 6.82
Current or Ex-Smoker	6	8	14

Table 1: Participant characteristics.

3. Results

- Sixty-five volatile compounds were included in the final curated data set and 14 were assigned a tentative identification based on matching to the National Institute of Standards and Technology (NIST) database (Table 2).
- Tetrachloroethylene was a top biomarker distinguishing CDI cases and controls by paired and group (Table 3) significance testing as well as by classification modeling (Table 3).
- Nine tentatively identified VOCs were selected by a quadratic discriminant analysis modeling method (Table 3). The model was able to distinguish between CDI cases and controls with 0.74 accuracy, sensitivity of 0.71, a specificity of 0.76, and a mean area under the receiver operating characteristic curve of 0.72 (Fig. 2).

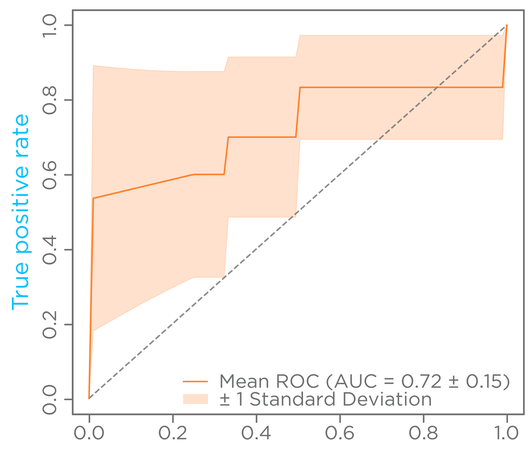

Figure 2: Receiver operating characteristic (ROC) curve for the QDA model. *A model with 9 VOCs had an accuracy 0.74, AUC 0.72, sensitivity 0.71, and specificity 0.76.*

MF	NIST Tentative ID	NIST % Match	Uncorrected *P*-value	Fold Change (case/control)
MF2	n-Hexane	81.93	0.023	1.343
MF18	Tetrachloroethylene	97.59	0.034	0.673
MF23	p-Xylene	93.8	0.039	1.594
MF55	Undecane, 3-Methyl-	89.8	0.043	1.230
MF53	2-phenyl-2-propanol	80.87	0.065	2.468
MF25	4-Heptanone	93.79	0.069	2.434
MF26	o-Xylene	84.98	0.069	1.476
MF16	Toluene	87.05	0.074	1.383
MF36	5-Hepten-2-one, 6-methyl-	95.58	0.114	0.711
MF41	1-Hexanol, 2-Ethyl	94.98	0.128	1.142
MF46	7-Octan-2-ol, 2,6-dimethyl-	78.71	0.128	3.255
MF56	Isophorone	75.75	0.143	1.108
MF31	Benzene, 1,2,4-trimethyl-	71.41	0.151	1.745
MF50	3-Octanol, 3,7-dimethyl-	91.29	0.176	1.201

Table 2: Fourteen volatile compounds with tentative identifications are shown along with their NIST scores and Mann-Whitney U-test P values. *n-Hexane (MF2), 3-methylundecanes (MF55) and 2-phenyl-2-propanol (MF53) are biomarkers with potential connections to lipid peroxidation and oxidative stress; 4-Heptanone (MF25) and 2-Ethyl-1-Hexanol (MF41) are VOCs with potential connections to phthalate plasticizers; multiple* Clostridium *species exhibit the capacity to grow in the presence (and perhaps degrade) tetrachloroethylene (MF18), but this metabolic capability has not been described for C. difficile[2]. Molecular features were assigned IDs in retention time order.*

MF	NIST Tentative ID	NIST % Match	Mann-Whitney *P*-value	Wilcoxon sign-rank *P*-value	Effect Size
MF2	n-Hexane	81.93	0.023	0.480	0.522
MF18	Tetrachloroethylene	97.59	0.034	0.099	-0.219
MF53	2-phenyl-2-propanol	80.87	0.065	0.308	0.682
MF26	o-Xylene	84.98	0.069	0.136	0.549
MF25	4-Heptanone	93.79	0.069	0.136	0.720
MF41	1-Hexanol, 2-Ethyl	94.98	0.128	0.050	0.132
MF46	7-Octan-2-ol, 2,6-dimethyl-	78.71	0.128	0.272	0.451
MF31	Benzene, 1,2,4-trimethyl-	71.41	0.151	0.480	0.266
MF50	3-Octanol, 3,7-dimethyl-	91.29	0.176	0.136	0.251

Table 3: Summary of the volatile compounds selected by the QDA classifier model. 21 MFs with P-value < 0.2 were considered for model building A non-linear quadratic discriminant analysis model delivered the best accuracy of the methods tested. QDA models with 9-11 VOCs had the greatest accuracy. A model with 9 VOCs features was selected to prevent the likelihood of overfitting.

4. Key Biomarker Highlights, Conclusions and Future Directions

- Lipid peroxidation, as potentially indicated by statistically significant increases of the 2 alkanes n-Hexane and 3-methylundecanes (Fig. 3), appeared to be higher in CDI patients than in controls.

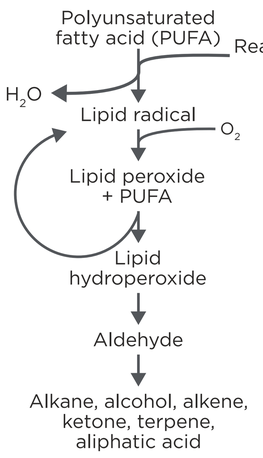

Figure 3: Lipid peroxidation metabolic pathway [3].

- 4-heptanone, a volatile compound linked to metabolic degradation of phthalate-containing plastics (Fig. 4), was elevated in the CDI group. Follow up work is needed to determine if the presence of 4-heptanone signifies that it is a novel risk factor for CDI or whether it reflects an artifactual in-hospital exposure to plastics. The potential connection of 4-heptanone and 2-ethyl-1-hexanol (and di-(2-ethylhexyl)-phthalate (DEHP) exposures) to CDI could be considered by measuring mono-(2-ethylhexyl)-phthalate-β-D-glucuronide in urine or blood, and examining medical records for the study population to look for differences in potential exposures to medical grade plastics.

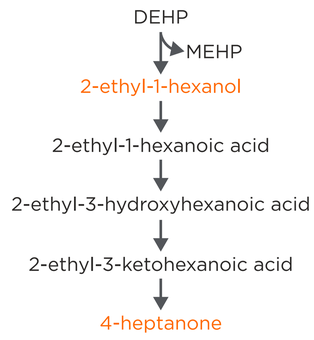

Figure 4: Phthalate metabolism leading to 4-heptanone production as described in [4].

- The significant decrease of tetrachloroethylene and the reported ability of some Clostridium species to metabolize tetrachloroethylene (Fig. 5) suggests that its relative decrease in CDI patients is a real finding related to its degradation by *C. difficile*. The ability of C. difficile to metabolize tetrachloroethylene could be examined in vitro to see if it can degrade this compound as a way to explain the relative decrease in the breath of the CDI patients.
- Nine of the named VOCs were selected by a quadratic discriminant analysis modeling method, were able to classify between CDI cases and controls 74% accuracy, sensitivity of 0.71, a specificity of 0.76, and a mean area under the curve (AUC) of 0.72.

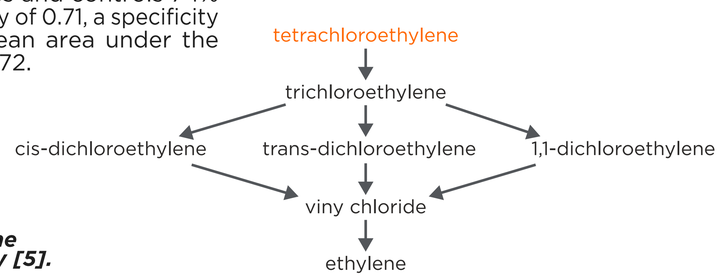

Figure 5: Tetrachloroethylene metabolic pathway [5].

5. References

1 Smits, W.K., et al., Clostridium difficile infection. Nat Rev Dis Primers, 2016. 2: p. 16020.

2 Bowman, K.S., F.A. Rainey, and W.M. Moe, Production of hydrogen by Clostridium species in the presence of chlorinated solvents. FEMS Microbiol Lett, 2009. 290(2): p. 188-94.

3 Kivits, G.A., M.A. Ganguli-Swarttouw, and E.J. Christ, The composition of alkanes in exhaled air of rats as a result of lipid peroxidation in vivo. Effects of dietary fatty acids, vitamin E and selenium. Biochim Biophys Acta, 1981. 665(3): p. 559-70.

4 Wahl, H.G., et al., 4-Heptanone is a metabolite of the plasticizer di(2-ethylhexyl) phthalate (DEHP) in haemodialysis patients. Nephrol Dial Transplant, 2004. 19(10): p. 2576-83.

5 Okeke, B.C., et al., Purification, cloning and sequencing of an enzyme mediating the reductive dechlorination of tetrachloroethylene (PCE) from Clostridium bifermentants DPH-1