

Application of Volatile Organic Compound analysis to differentiate between bacterial and viral respiratory infections

A thesis submitted to Imperial College for the degree of Doctor of Philosophy by

Dr Faisal Kamal

Department of Surgery and Cancer,
Imperial College London,
St Mary's Hospital,
London, UK

2019

Supervisors

Professor George B Hanna PhD FRCS

Department of Surgery and Cancer, Imperial College London,
St Mary's Hospital, London, UK

Professor Sebastian Johnston PhD FRCP

National Heart & Lung Institute, Imperial College London,
St Mary's Hospital, London, UK

Declaration of originality

All the work presented in this thesis was carried out by myself, as part of my PhD work at Imperial College London, unless stated and referenced in the text. Mr Pranav Patel corroborated the results of my literature review by checking I had not missed any papers in my search.

I carried out measurement of Volatile Organic Compound from the headspace of agar plates cultured with bacteria as well as the Bronchial Epithelial Cell (BEAS2B) cultures infected with viral and bacterial infections. These experiments were set up with the assistance of Dr Mike Edwards and Dr Aran Singanayagam. All VOC collection samples I ran myself on the Gas Chromatography Mass Spectrometry spectrometer Agilent instrument.

Regarding the human RV challenge, I was present for nearly every screening and sampling visit, with a small number covered by a research nurse, Belen Trujillo-Torralbo. All breath collections I undertook and ran on GC-MS instruments myself, with subjects organised by Dr Ross Walton and Mike Edwards who were involved in the organisational and administrative aspects of this study.

The London COPD prospective study was already well established and I joined them in November of 2016 to collect breath from patients. The cohort has had varying levels of additional assistance provided by one appointed nursing staff and 3 other clinical research fellows. I helped in the cohort follow up by providing my services twice a week, running what were essentially clinics, to see both stable and acutely exacerbating patients. The majority of breath samples were collected by myself, with my colleagues providing assistance during days on which I was unable to attend clinic. All breath samples were transported from Royal Brompton Hospital and run on two instruments by myself. Sputum microbiology analysis was undertaken by Royal Brompton hospital laboratories and sputum viral PCR analysis was undertaken by Dr Tata Keadze at Imperial College, St Mary's campus. License terms of this work.

Copyright declaration

'The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.'

Acknowledgements

I would like to express my deepest gratitude to the many people, who have helped me during the course of my research. The first person I would like to thank is my primary supervisor, Professor George B Hanna. He provided me with the opportunity to take time out of my clinical training and, with his guidance and support, has been instrumental in helping me to complete this research. Professor Hanna has been an inspirational role model over the past three years and I have learnt many important lessons which will remain with me throughout my career. I am indebted to Professor Sebastian Johnston and Professor Wisia Wedzicha, both of whom provided me with invaluable advice and their expert insights during my research. I have been genuinely lucky to have their continuous support through both of their laboratories and research groups. This was essential to the completion of this thesis, in addition to providing me with some new, lifelong friends in colleagues that I met within their respective groups.

My other supervisors, Mr Sacheen Kumar and Dr Aran Singanayagam, provided an unparalleled day to day level of support. Their mentorship, support in the laboratory and friendship throughout the last three years has been incredible and I am most grateful.

I would like to thank Dr Andrea Romano who was an ever present in the laboratory, helping with key analytical work. This was later supported, on their arrivals, by Ilaria Belluomo and ST Chin. I would like to thank Dr Mike Edwards and Ross Walton for all their help but, in particular, the opportunity to collect breath from rhinovirus infected healthy human subjects (help with patient recruitment, organisation and in the collecting of clinical data). Mike and Aran were ever present laboratory partners, supportive and present at all stages of my *in vitro* experiments. I am also grateful to Dr Kirill Veselkov and Dr Ivan Laponogov for their bioinformatics help with statistical analysis.

My final words of gratitude and appreciation are for my wife Laura and my family; in particular my parents. I am, and always will be, eternally grateful for the unwavering support you have all given me. Laura, my best friend, thank you for your patience, love and support through this challenging and exciting time.

Abstract

Respiratory infections are a leading cause of morbidity and mortality. Bacteria and viruses precipitate the majority of respiratory tract infections and also induce exacerbations in patients. Infection aetiology is particularly difficult to distinguish clinically and this represents a major unmet clinical need. Accurate identification of infection aetiology of respiratory infections would facilitate initiation of the most appropriate prompt management.

There is evidence from *in vitro* and *in vivo* models that Volatile Organic Compounds (VOCs) can be measured through multiple analytical mass spectrometry platforms to differentiate between pathogens.

The aim of this PhD is to investigate the utility of VOC measurements as a method of identification of infection aetiology and to determine whether such tests could be useful in differentiation of bacterial and viral exacerbations. Gas Chromatography-Mass Spectrometry (GC-MS) will be the primary method of analysis.

Hypothesis

Interaction between bacteria or viruses with their host results in different biochemical reactions. Metabolic degradation products including VOCs can be detected in exhaled breath and can be utilised to identify aetiology of exacerbations.

Sub hypothesis

- Pathogenic bacteria produce distinct VOCs that can be measured in the headspace following culture on agar medium.
- Infection of airway epithelial cell cultures *in vitro* with pathogenic bacteria and respiratory viruses will lead to production of virus and bacteria specific VOCs that can be detected in the headspace.
- Exhaled breath sampling following experimental infection with rhinovirus 16 in healthy subjects will identify VOCs that are upregulated following virus infection *in vivo*.
- Targeted breath analysis in COPD patients during acute exacerbation will allow identification of pathogen-specific VOCs identified in sub-hypotheses 1, 2 and 3

Results

In vitro agar plate growth of *Streptococcus pneumoniae* (SP) and *Haemophilus influenzae* (HI) showed different VOC profiles which allowed for differentiation of the two bacteria species. The higher the concentration of streptococcus plated, there was noted to be a dose dependent increase in VOCs detected.

23 and 13 VOCs respectively were produced by SP and HI culture on agar plates. 8/23 VOCs showed dose dependent increase with bacterial concentration plated. Similarly, VOCs specific to viral and bacterial infection were identified from bronchial epithelial cells stimulated with rhinovirus (RV) 16, respiratory syncytial virus (RSV), SP and HI. Bacterial markers *in vitro* were able to be separated with an area under curve (AUC) of 0.89.

Long chain alkanes, such as decane were found in the exhaled breath of healthy subjects following RV infection. Findings from breath collection from COPD patients identified VOCs to differentiate between subjects who were having exacerbation compared to baseline state; subjects with either viral or bacterial positive sputum in exacerbations compared to those at baseline with negative microbiology could also be differentiated (later AUC 0.74).

Conclusion

Specific VOCs are associated with viral and bacterial infection. VOCs could provide an exciting tool to accurately determine aetiology, guiding judicious antimicrobial use.

Abbreviations

AMR	Antimicrobial resistance
BEAS-2B	Immortalised human Bronchial epithelial cells
CAP	Community acquired pneumonia
COPD	Chronic Obstructive Pulmonary Disease
CFU	Colony forming units
ELISA	Enzyme-linked immunosorbent assay
FEV1	Forced expiratory volume in 1 second
FBC	Full blood count
FCS	Foetal calf serum
FDR	False discovery rate
GOLD	Global initiative for chronic obstructive lung disease
GC-MS	Gas Chromatography-Mass Spectrometry
HI	<i>Haemophilus influenza</i>
KW	Kruskal-Wallis
IL	Interleukin
ILD	Interstitial lung disease
IP 10	Interferon-gamma inducible protein 10kDa/CXCL10
M	Media
MALDI -TOF-MS	Matrix-assisted laser desorption/ionization-time of flight-mass spectrometry
MC	Media and cells (BEAS2b)
MOI	Multiplicity of Infection
NICE	National Institute for Health and Care Excellence
NIST	National institute of standards and technology
PBS	Phosphate Buffer solution
PCR	Polymerase chain reaction
PEF	Peak expiratory flow
PTR-TOF MS	Proton transfer reaction-Time of flight-Mass spectrometry
RT	Retention Time
RSV	Respiratory syncytial virus
RV	Rhinovirus
SP	<i>Streptococcus pneumoniae</i>
SVM	Support Vector Machine
TD	Thermal desorption
TH	Todd Hewitt
UV-RV	UV inactivated RV
UV-RSV	UV inactivated RSV

Abstracts and presentations of work contained in thesis:

Some of the data reported in this thesis has been previously presented at the following meetings:

1. F. Kamal, S. Kumar, A. Singanayagam, M. Edwards, A. Romano, K. Veselkov, I Laponogov, G. Donaldson, S. Johnston, J Wedzicha, G. Hanna. Volatile Organic Compound analysis to differentiate between bacterial and viral respiratory infections in COPD. Oral presentation at Division of Surgery Cancer, Faculty of Medicine, Winter meeting 2017.
2. F. Kamal, S. Kumar, A. Singanayagam, M. Edwards, A. Romano, K. Veselkov, I Laponogov, G. Donaldson, S. Johnston, J Wedzicha, G. Hanna. Volatile Organic Compound analysis to differentiate between bacterial and viral respiratory infections in COPD. Presentation European Respiratory Society, Paris, Wednesday 19th September 2018.
3. F. Kamal, S. Kumar, A. Singanayagam, M. Edwards, A. Romano, K. Veselkov, I Laponogov, G. Donaldson, S. Johnston, J Wedzicha, G. Hanna. Volatile Organic Compound analysis to differentiate between bacterial and viral respiratory infections. Oral Presentation at the Merieux research grant meeting. Annecy France. 30th-31st October 2018.

Table of contents

1	Introduction	16
1.1	Respiratory infections.....	16
1.1.1	Respiratory tract Infections in otherwise healthy individuals	16
1.2	Chronic obstructive pulmonary disease (COPD)	17
1.2.1	Clinical burden of the disease	17
1.2.2	Incidence, Aetiology & Symptomatology.....	17
1.2.3	Exacerbations	18
1.2.4	Viruses as a trigger for exacerbation	18
1.2.5	Bacteria and exacerbations.....	18
1.2.6	Dual virus/bacterial infections	19
1.3	Currently available approaches to differentiating between virus and bacteria-induced exacerbations	20
1.4	Adverse consequences of antibiotic over-use.....	21
1.5	Overview of the use of Volatile Organic Compound measurement in clinical practice	21
1.6	Overview of Mass Spectrometry	23
1.7	Mass Spectrometry techniques.....	23
1.7.1	Gas Chromatography Mass Spectrometry	23
1.7.2	Limitations of GC-MS	26
1.8	Overview of VOCs in respiratory disease	26
1.9	Use of mass spectrometry in bacterial pathogen identification	27
1.10	Use of mass spectrometry in viral pathogen identification	28
1.11	Existing studies that evaluate exhaled breath to determine exacerbation aetiology.....	28
1.12	Aim and Hypotheses.....	30
1.12.1	Overall Hypothesis:	30
1.12.2	Sub Hypotheses:.....	30
2	Review VOCs in common community infections	31
2.1	Introduction	31
2.2	Methods.....	32
2.3	Results of literature search.....	36

2.3.1	Streptococcus pneumoniae:	36
2.3.2	Haemophilus influenzae.....	37
2.3.3	Influenza and rhinovirus	37
2.4	Discussion	38
3	Evaluation of VOC production by bacteria cultured on agar plates	40
3.1	Objectives	40
3.2	Introduction	40
3.3	Material & Methods	40
3.3.1	Plating and culturing bacteria	40
3.3.2	VOC collection	41
3.3.3	TD tube preparation.....	42
3.3.4	GC-MS analysis.....	42
3.3.5	Data analysis	43
3.4	Results.....	44
3.4.1	<i>S. pneumoniae</i> growth on agar plates	44
3.4.1.1	Comparison of effects of Tedlar bag encasement on bacterial growth	44
3.4.1.2	Comparison of headspace VOCs between high dose <i>S. pneumoniae</i> versus media and uninfected plate control on chocolate agar at 24h	44
3.4.1.3	Effect of infective dose on VOC production in headspace	46
3.4.2	<i>H. influenzae</i> growth on agar plates	47
3.4.3	Species specificity and overlap between <i>S. pneumoniae</i> and <i>H. influenzae</i> VOC production.....	48
3.5	Discussion	48
3.6	Conclusion	51
4	Cell culture, infection and VOC collection	52
4.1	Objectives	52
4.2	Introduction	52
4.3	Methods.....	52
4.3.1	Cell Culture.....	53
4.3.2	Rhinovirus infection	53
4.3.3	RSV infection	53
4.3.4	Bacterial infection	53
4.3.5	VOC collection	54
4.3.6	GC-MS analysis.....	54
4.3.6.1	Standards	55

4.4	Results.....	55
4.4.1	Evaluation of effect of Tedlar bag enclosure on BEAS2B cell growth.....	55
4.4.2	Comparison of VOC production from BEAS2B cells in medium versus medium alone	56
4.4.3	RV16 infection in BEAS2B airway epithelial cells.....	56
4.4.3.1	Cytokine responses	56
4.4.3.2	Measurement of VOCs in headspace of RV infected cells	57
4.4.4	RSV infection in BEAS2B cells	63
4.4.4.1	Cytokine responses	63
4.4.4.2	Evaluation of VOC production in headspace of RSV infected cells	64
4.4.4.3	Batch adjusted analysis RSV infection	68
4.4.5	Overlap VOCs between RV and RSV	70
4.4.6	Measurement of VOCs in headspace of <i>S. pneumoniae</i> and <i>H. influenzae</i> infected cells	71
4.4.7	Combined bacterial infection analysis	73
4.5	Discussion	79
4.6	Conclusion	81
5	VOC collection from human clinical studies of healthy subjects	83
5.1	Introduction.....	83
5.2	Method	84
5.2.1	Study Design.....	84
5.2.2	Sampling and assessments.....	86
5.2.3	Experimental RV16 infection	87
5.2.4	Measurements of virus load	87
5.2.5	VOC transfer to thermal desorption tube.....	87
5.2.6	GC-MS analysis	88
5.2.7	Data analysis	89
5.3	Results.....	90
5.3.1	Assessments of virus load	90
5.3.2	Identification of VOCs in infected subjects	91
5.3.3	Association of VOCs and virus load.....	93
5.3.4	Associations of VOCs and Symptom score	95
5.4	Discussion	98
5.5	Conclusion	100
6	London COPD cohort	101
6.1	Introduction	101

6.2 Method	102
6.2.1 Study design	102
6.2.2 VOC breath collection method.....	103
6.2.3 VOC mass spectrometry analysis	103
6.2.4 Samples other than breath collected.....	104
6.2.5 Data analysis	104
6.3 Results	106
6.3.1 Overview of samples collected	106
6.3.2 Microbiological and virological assessments in sputum	107
6.3.3 Analysis of VOCs during exacerbation compared to baseline	108
6.3.4 Analysis of VOCs during exacerbation, baseline and recovery	110
6.3.5 Temporal analysis of VOCs	112
6.3.6 Analysis of VOCs in Rhinovirus positive subjects compared to rhinovirus negative subjects (irrespective of clinical stability)	113
6.3.7 Analysis of VOCs during RV exacerbation (RV positive, bacterial negative) compared to baseline (virus and bacterial negative) breath.....	115
6.3.8 Analysis of VOCs in RV positive exacerbation (bacterial negative) compared to RV negative exacerbation (bacterial negative)	117
6.3.9 Analysis of VOCs in RSV positive subjects compared to RSV negative (all clinical states)	118
6.3.10 VOC Analysis during Bacterial colonisation: Baseline positive subjects compared to baseline negative subjects	120
6.3.11 Analysis of VOCS during Bacterial positive exacerbations (virus negative) Vs Baseline (negative virus and bacteria)	122
6.3.12 Analysis of VOCs in <i>Streptococcus pneumoniae</i> and <i>Haemophilus influenza</i> positive samples at exacerbation compared to baseline (bacteria and virus negative)	126
6.3.13 Analysis of VOCs in exacerbation and baseline samples using a Paired analysis	128
6.3.14 Analysis of VOCs using regression analysis	128
6.4 Discussion	130
6.5 Conclusion	134
7 Summary, conclusion and future direction	135
7.1 Summary and Conclusion	135
7.2 Future directions	141
8 Appendices	143
8.1 Cohort demographics and Inclusion and exclusion criteria	143
8.2 Sputum processing	148
8.3 Viral RNA extraction	149

8.4	COPD full data tables	154
8.5	Metabolic Pathways	171
9	References	173

List of Figures

Figure 1: TD GC-MS schematic.....	25
Figure 2: Systematic review overview VOCs and common respiratory pathogens.....	33
Figure 3: VOCs collection model from agar plates.....	42
Figure 4: VOC induction by <i>S. pneumoniae</i> noted on growth in agar plates.....	45
Figure 5: Dose dependent VOC induction by <i>S. pneumoniae</i> grown in agar plates....	46
Figure 6: VOC induction by <i>H. influenzae</i>	47
Figure 7: VOC induction by <i>S. pneumoniae</i> and <i>H. influenzae</i>	48
Figure 8: VOC collection from cell model.....	54
Figure 9: Viability of BEAS-2B cells grown in a Tedlar bag.....	56
Figure 10: BEAS-2b cells inflammatory cytokine release in RV infection.....	57
Figure 11: ROC curve separation between control and cells infected with RV.....	58
Figure 12: VOC induction of Peak 179 in headspace of rhinovirus infected cells	60
Figure 13: Figure showing decane levels comparing RV infected cells to control.....	61
Figure 14: 2 methyl 2 propanol levels comparing RV infected cells to control.....	62
Figure 15: Chromatogram generated by standards of decane and dodecane.....	62
Figure 16: Inflammatory cytokine upregulation in RSV infection of BEAS2b.....	63
Figure 17: Output from MATLAB analysis.....	66
Figure 18: 3-pentanone levels comparing RSV infected cells to control.....	67
Figure 19: Acetic acid (acetyloxy) comparing RSV infected cells to control.....	67
Figure 20: Isopropyl lactate levels comparing RSV infected cells to control.....	70
Figure 21: Decanal levels comparing RSV infected cells to control.....	70
Figure 22: Venn diagram comparing VOCs in RV and RSV BEAS2b infection.....	71
Figure 23: Neopentane levels comparing bacteria infected cells to control.....	72
Figure 24: Dimethyl disulfide levels comparing bacteria infected cells to control.....	73
Figure 25: Separation between control and cells infected with bacteria.....	74
Figure 26: Induction of dimethyl disulfide in cells infected with bacteria	75
Figure 27: Methyl nitrite levels in bacteria infected cells comparing control	75
Figure 28: Separation between control (media) and media containing bacteria.....	76
Figure 29: 4,5 dimethyl1,3-dioxl-2-one levels in bacteria and media and media.....	78
Figure 30: Neopentane levels in bacteria and media (no cells) comparing media.....	78
Figure 31: Venn diagram summarising findings.....	80
Figure 32: Overall schematic for RV challenge trial in human volunteers.....	85

Figure 33: Human breath collection.....	88
Figure 34: Mean viral loads in nasal passage and oropharynx.....	90
Figure 35: VOC induction in RV16 infection of healthy human 2 of 12 VOCs.....	91
Figure 36: 14 cluster patterns in human RV infection.....	92
Figure 37: Regression analysis relationship between nasal viral load and VOCs.....	94
Figure 38: Benzyl methyl sulphide regression analysis.....	95
Figure 39: Showing mean of total symptom scores.....	96
Figure 40: Relationships between peak VOC and peak symptom score.....	97
Figure 41: Relationships between peak VOC and total symptom score.....	98
Figure 42: Study design for investigating VOCs in exacerbations of COPD.....	103
Figure 43: Overview of samples collected and analysed.....	107
Figure 44: 2,9-dimethyl-undecane higher in exacerbation breath samples	108
Figure 45: 1,4 pentadiene significantly higher in exacerbation breath samples	109
Figure 46: Cyclopentanone significantly higher in exacerbation breath samples.....	109
Figure 47: 2 ethyl hexane levels from baseline, acute exacerbation and recovery...111	
Figure 48: 1,4 pentadiene levels from baseline, acute exacerbation and recovery...111	
Figure 49: 2 ethyl hexane levels from baseline, acute exacerbation and recovery....112	
Figure 50: 1,4 pentadiene levels baseline through acute exacerbation recovery....113	
Figure 51: Peak 667 increase RV viral infections but not identifiable through NIST..114	
Figure 52: Peak 720 increase RV viral infections but not identifiable through NIST...116	
Figure 53: 2,6,10-trimethyl-Dodecane increased in viral infection.....	116
Figure 54: Peak 751, rt 29.00, likely decane in RV COPD exacerbation.....	117
Figure 55: Peak 591, rt 22.57 m/z base peak 45 increase in RSV COPD infection...119	
Figure 56: Peak 322, rt 13.94 m/z base peak 42 increase in RSV COPD infection...120	
Figure 57: Methyl methacrylate levels increased in bacterial colonisation.....	122
Figure 58: ROC AUC curve for bacterial exacerbation and healthy subjects.....	123
Figure 59: Acetic acid levels in bacterial exacerbation breath samples.....	124
Figure 60: Decrease in propanoic acid in exacerbation breath samples.....	124
Figure 61: ROC curve for SP and HI bacterial exacerbation breath samples.....	127
Figure 62: ROC AUC curve exacerbation breath samples compared to healthy subjects breath and regressed for patient factors.....	130

List of Tables

Table 1: A summary of a systematic review of the literature describing VOCs produced from common respiratory infections.....	34
Table 2: Results comparing my results to previous literature.....	50
Table 3: Buffer and reagents.....	52
Table 4: Table shows 14 discriminatory identified through ANOVA and FDR correction, RV infected cells versus control (media and cells).....	59
Table 5: Table shows 14 discriminatory identified through ANOVA and FDR correction, RSV infected cells versus control (media and cells).....	65
Table 6: Table showing discriminatory VOCs in RSV infected cells versus control (media and cells). Medium and cell baseline corrected with time adjustment.....	68
Table 7: Discriminatory VOCs between <i>S. pneumoniae</i> and <i>H. influenzae</i> and cell versus control (media and cells).....	72
Table 8: Discriminatory VOCs between combined bacteria <i>S. pneumoniae</i> and <i>H. influenzae</i> and cell versus control (media and cells).....	74
Table 9: Discriminatory VOCs between combined bacteria <i>S. pneumoniae</i> and <i>H. influenzae</i> and media versus media alone (media no cells).....	77
Table 10: Top 15 VOCs increasing in RV human challenge.....	93
Table 11: VOCs significantly changed during exacerbation comparing baseline breath samples.....	110
Table 12: VOCs changes in RV positive breath comparing RV negative breath.....	114
Table 13: VOCs changes comparing RV positive breath samples versus baseline breath samples (negative bacteria and virus).....	117
Table 14: VOCs changes comparing RV positive exacerbation breath samples versus RV negative exacerbation samples.....	118
Table 15: VOCs changes comparing RSV positive breath samples versus RV negative samples.....	119
Table 16: VOCs changes comparing bacterial positive breath samples at baseline versus URTF at baseline samples.....	121
Table 17: VOCs changes comparing bacterial positive breath samples at baseline versus baseline (negative bacteria and virus) breath samples.....	125
Table 18: VOCs changes comparing bacterial (<i>S. pneumoniae</i> and <i>H. influenzae</i> only) positive breath samples at baseline versus baseline (negative bacteria and virus) breath samples.....	127
Table 19: Top 15 VOCs changes comparing exacerbation breath samples versus baseline (negative bacteria and virus) breath samples undertaking paired analysis.....	128
Table 20: Smoking related VOCs changes identified comparing smoking breath samples versus non-smoking breath samples.....	129

1 Introduction

1.1 Respiratory infections

1.1.1 Respiratory tract infections in otherwise healthy individuals

Respiratory infections are a leading worldwide cause of mortality and morbidity¹. Some infections such as rhinovirus cause mild self-limiting illness in healthy subjects whilst others, such as the pathogenic bacteria *Streptococcus pneumoniae* (*S. pneumoniae*) and *Haemophilus influenzae* (*H. influenzae*), can precipitate more severe illness in the form of community-acquired pneumonia².

Human rhinovirus (RV) is the most common cause of upper respiratory tract infection (URI) and development of the common cold³. Despite the relative self-limiting nature of RV infection in the majority of healthy individuals, it is still a significant economic burden on society. It leads to increased consultations with clinicians, increased absence from school and work and a subsequent loss in earnings^{4,5}.

Community acquired pneumonia (CAP) has a high (incidence 1.7-11.6 cases per 1000 people per year)¹, and is the most frequent respiratory cause of death in Europe. *S. pneumoniae* has been identified as the most common causative pathogen⁶. Pneumonia disproportionately affects older people⁷. The reported incidence of invasive pneumococcal disease (IPD) in the UK is 6.85 per 100,000 annually⁸. Pneumonia is the sixth biggest cause of death in the UK and of ~220,000 diagnosed per year, ~29,000 (13%) die from the disease⁸. Current guidelines advocate use of clinical severity scores guiding antibiotic prescribing but recent studies show this leads to inadequate treatment⁹. This approach has theoretical advantages because a decrease in broad-spectrum antibiotic treatment in low-risk patients might reduce antibiotic-related side-effects, and to give broad-spectrum therapy to patients at higher risk of death⁹. Use of markers such as calcitonin to help guide antibiotic therapy have been trialled with both favourable¹¹ and unfavourable¹² antibiotic reduction results. Procalcitonin has been seen to be more sensitive than c reactive protein to indicate bacterial infection but remains recommended not be used in isolation as a rule-out tool¹³.

The mortality rate remains high and new diagnostic tests to determine aetiology at the front door are needed. The development of breath diagnostics, as an adjunct to current methods to improve diagnostics, could thus drive a further improvement in clinical care through more accurate detection and definitive treatment.

1.2 Chronic obstructive pulmonary disease (COPD)

1.2.1 Clinical burden of the disease

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality worldwide¹⁴. The Global Initiative for Obstructive Lung disease (GOLD) guidelines define COPD as a common preventable disease that is characterised by obstructive airflow limitation, which is normally progressive over time¹⁵. The persistent airflow limitation results from a combination of diffuse small airway disease and destruction of lung parenchyma (emphysema)¹⁶.

1.2.2 Incidence, Aetiology & Symptomatology

In the UK, approximately 1.2 million people (~2-3 % of the population) have been diagnosed with COPD¹⁷. An estimated 30,000 people die of the disease in the UK every year¹⁷; it is also a major cause of morbidity and mortality worldwide¹⁸. The disease is associated with a chronic inflammatory process induced by cigarette smoke, the most important and modifiable aetiological factor for COPD¹⁹. Smoking cessation and importance of smoking cessation programmes with regards to improving long term health outcomes has been repeatedly evidenced²⁰. Other important risk factors include: occupational airborne pollutants, genetic factors (alpha 1 antitrypsin deficiency), air pollution, poor socioeconomic status, early life exposure to cigarette smoke and recurrent childhood infections¹⁶. In COPD, the most important chronic clinical symptoms include breathlessness and cough¹⁶. The severity of COPD is classified based on a number of factors including airway obstruction, exacerbation frequency and degree of dyspnoea/limitation²¹.

Studies by multiple groups demonstrate that the lower airways are not sterile and the existence of a bacterial microbiome^{22,23}. The importance of the role lung microbiota and host response at all stages of COPD have been highlighted²². A consistent finding across studies is an increase of the Proteobacteria phylum and a decrease in the

Bacteroidetes phylum in late-stage COPD vs control subjects²⁴. Observations from bacterial microbiome research indicate that lung communities may be involved in the progression of COPD and that microbes identified in the lung are associated with host inflammation^{25,26}. The imbalance of bacteria and subsequent inflammation has led to the proposed microbiome dysbiosis-inflammation cycle thought to lead to exacerbations and COPD progression^{26,27}.

1.2.3 Exacerbations

The GOLD strategy defines an exacerbation of COPD as “an event in the natural course of the disease characterized by a change in the patient’s baseline dyspnoea, cough, and/or sputum that is beyond normal day to-day variations, is acute in onset, and may warrant a change in regular medication in a patient with underlying COPD”²⁸. Exacerbations can occur at any stage in the disease course²⁹ and may have both short and long term clinical implications. These include: requirement for hospitalisation, accelerated lung function decline³⁰, impaired quality of life³¹, increased risk of death³² and healthcare related- costs¹.

1.2.4 Viruses as a trigger for exacerbation

Around 70-80% of COPD exacerbations are associated with viral respiratory infections³³. The most commonly detected viruses at exacerbation are rhinoviruses (RV)³⁴. Other implicated viruses include: respiratory syncytial virus (RSV), influenza and coronaviruses³⁵. Studies carrying out experimental virus challenge in subjects with COPD have been pivotal in demonstrating a causal relationship between respiratory virus infection and exacerbation¹⁵. In COPD, experimental RV challenge is associated with greater inflammation, increased symptom severity/duration and greater acute change in lung function compared to healthy subjects³⁶.

1.2.5 Bacteria and exacerbations

Bacteria can cause chronic infection/colonisation and are also associated with acute exacerbations in COPD. The most common organisms cultured in COPD at stable state, from both sputum and bronchoscopic samples, are consistently *H. influenzae*, *S. pneumoniae*, and *Moraxella catarrhalis*^{37,38}. The spectrum of colonising pathogenic bacteria varies with disease severity with gram negative organisms, such as *Pseudomonas aeruginosa*, being more frequent in those with severe airflow obstruction³⁷.

The role of bacteria in causing exacerbations is unclear. Studies by Hurst *et al*³⁹ and Wilkinson *et al*⁴⁰ observed increased bacterial load at exacerbation compared to stable state. Conversely, Sethi *et al* showed no difference in concentrations of *H. influenzae* or *Moraxella catarrhalis* from stable state to exacerbation³⁷ but did show that new bacterial strain acquisition is associated with development of exacerbations⁴¹.

Modern molecular techniques have revolutionised our understanding of bacteria in COPD and revealed the existence of complex communities of commensals within the microbiota⁴². Wang *et al* demonstrated that the lung microbiota changed during acute exacerbation events and described dynamic changes specific to exacerbation phenotypes⁴³. They observed reduced overall diversity with a small nonsignificant microbial composition shift toward an increase in the relative abundance of Proteobacteria (adj. $p=0.42$ using paired t-test) and a decrease in Firmicutes (adj. $p=0.73$ using paired t-test) during exacerbations compared to stable samples. *Moraxella* showed the greatest change during exacerbation compared to stable ($p=0.022$ using t-test)⁴³.

1.2.6 Dual virus/bacterial infections

As well as viruses and bacteria individually triggering exacerbations, they can often co-exist during these episodes³⁷. The concept of virus-induced secondary bacterial infection has been raised by studies of naturally occurring exacerbations⁴⁴ and experimental challenge studies⁴⁵. In a study by Mallia *et al*, 60% of patients with COPD developed a secondary bacterial infection following experimental RV challenge, suggesting a causal link between viral and secondary bacterial infection⁴⁵. In this study, reduced airway levels of the antimicrobial peptides elafin, and secretory leucoprotease inhibitor (SLPI), mediated by virus-induced neutrophil elastase cleavage, were associated with increased risk of secondary bacterial infections⁴⁵. George *et al* studied patients from the London COPD cohort and identified that 73% of those initially bacterial negative and RV positive were bacterium positive from day 14⁴⁴, further supporting the phenomenon of virus-induced secondary bacterial infection. Like other types of bacterial pneumonia, pneumococcal pneumonia is especially common following viral infections⁴⁶. Viral infections lead to over-expression of pneumococcal binding receptors, impaired alveolar macrophage phagocytosis and neutrophil dysfunction^{47,48}. Sustained desensitization of lung sentinel cells to Toll-like

receptor (TLR) ligands lasts for several months after resolution of influenza or respiratory syncytial virus infection and is associated with reduced chemokine production and NF-kappaB activation in alveolar macrophages. Subsequent reduced neutrophil recruitment correlates with heightened bacterial load during secondary respiratory infection⁴⁸. This co-existence, or consequential occurrence of virus and bacterial infection, adds further complexity to the decision to treat exacerbations with antibiotics and thus cover bacterial infection. Therefore, at present, antibiotic prescribing for COPD exacerbations has to be empirical ¹.

1.3 Currently available approaches to differentiating between virus and bacteria-induced exacerbations

Bacterial and viral infections are not easily differentiated on initial presentation of a patient with an acute exacerbation of COPD. There is a lack of accurate diagnostic tools to diagnose aetiology of respiratory tract infections and this invariably leads to overtreatment with antibiotics. Despite the fact that the role of bacteria is unproven and that methods to identify bacteria as causative triggers for exacerbations are limited, antibiotic use for these episodes remains widespread (>80% in secondary care and ~50% in primary care)⁴⁹.

Both viruses and bacteria can induce similar clinical symptoms of cough, dyspnoea or increased sputum purulence, making clinical history an imprecise tool for determining aetiology⁵⁰. The routine microbiological tests undertaken on sputum (microscopy, culture and sensitivity), during acute admissions, do not allow rapid accurate identification of bacteria and can take up to several days to yield useful information. The decision to treat with antibiotics (and cover bacterial infection) at exacerbation therefore has to be based on imperfect surrogate clinical approaches, such as the Anthonisen criteria which comprises three patient-reported items: increased dyspnoea, increased sputum volume, and increased sputum purulence (a symptom complex used to indicate likelihood of bacterial infection)⁵¹. Therefore, there are currently no rapid clinical evaluation measures that facilitate accurate targeted antibiotic therapy at the front door. A biomarker, or panel of biomarkers, that would facilitate a precise aetiological diagnosis is urgently needed to enable appropriate targeted management. Previous studies have shown that systemic biomarkers of bacterial infection such as procalcitonin can be used to safely reduce antibiotic use but remain insensitive and measure the systemic response to infection and illness rather

than lung-specific processes^{52,53}. Bafadhel *et al* identified sputum IL-1 β as a potential marker for determining bacteria-associated exacerbation and serum CXCL10 (IP-10) has been proposed as a possible biomarker for rhinovirus infection at exacerbation⁵⁴. However, these tests are not specific and a diagnostic test that specifically identifies bacterial or viral respiratory infection is, to date, unavailable.

1.4 Adverse consequences of antibiotic over-use

Antibiotic resistance is a major worldwide public health problem, and international efforts are needed to counteract its emergence. Antibiotic consumption is increasingly being recognised as the main cause of this emerging resistance⁵⁵. Inappropriate over-use of antibiotics is associated with high healthcare costs⁵⁶ and contributes to the emerging problem of development of resistant pathogens and antibiotic-related complications such as *C. difficile* infections^{56,57}. The lack of an appropriate biomarker for bacterial respiratory infections has led to widespread over-prescription of antibiotics during exacerbations⁵⁸. Antimicrobial resistance (AMR) threatens the effective prevention and treatment of an ever-increasing range of infections and is a serious threat to global public health, which requires action across all government sectors and society⁵⁹. In Europe, AMR has been observed in all pathogens associated with CAP, including the commonly encountered *S. pneumoniae*⁶⁰. In a pan-European study, the overall percentages of isolates demonstrating intermediate or complete resistance to penicillin, erythromycin, tetracycline, trimethoprim-sulfamethoxazole (TMP-SMX) and ciprofloxacin were: 24, 24.6, 19.8, 26.7 and 2%, respectively⁶¹.

1.5 Overview of the use of Volatile Organic Compound measurement in clinical practice

Breath, and changes in its composition, has been noted in systemic disease states since the days of Hippocrates⁶²⁻⁶⁴. Human breath has been found to contain over a thousand constituents known as Volatile Organic Compounds (VOCs)⁶⁵. VOCs are a large group of carbon-based chemicals that can easily evaporate at room temperature. Historically, examination of the breath of patients has provided clinicians with important information to help diagnose underlying systemic diseases, e.g. foetor hepaticus and fetor oris in the context of chronic liver disease⁶⁶. Further evidence that analysis of breath could be extremely useful has been observed through canine olfaction, for identifying patients suffering from various diseases such as cancer⁶⁷, diabetic

ketoacidosis⁶⁸, hypoglycaemia⁶⁹ and epilepsy⁷⁰. Breath tests have now been approved for clinical use, such as exhaled nitric oxide to identify type 2 inflammation in asthma⁷¹ and breath tests to identify *Helicobacter pylori* in gastrointestinal disease⁷².

GC-MS instruments provide researchers and clinicians with a vast amount of information on sample composition, thus these instruments are seen as gold standard in breath analysis research⁷³. Much has been discussed about the difficulties of using GC-MS instruments for breath analysis in areas of standardisation, different instruments, collection materials, etc. Breath gas analysis is difficult to perform because of low VOC concentrations in breath air, high humidity of breath gas, lack of suitable sampling techniques, and absence of sensitive and capable measuring techniques.⁷⁴

Much of these difficulties have partly been overcome by improvements in sampling and analytical techniques. Preconcentrating breath gas by various means and subsequent analysis by means of gas chromatography combined with mass spectrometry (GCMS) constitute a reliable and sensitive set of methods for VOC analysis⁷⁵. Efforts to overcome these difficulties, both at the experimental and computational level, have become a focus of metabolomic research⁷⁶. These efforts include research on standardizing the experimental protocols⁷⁷ and strategies for incorporating quality controls⁷⁸. Metabolite identification is based on retention index and spectral match scores. Identified (known) analytes with reverse scores greater than or equal to 750, included in further analysis⁷⁹. The high sensitivity of GC-MS leads to the difficulty of distinguishing contaminants from potential VOCs of metabolic significance. Intra-instrument variability, variabilities of sample collection setup, environmental conditions (temperature, humidity etc), patient-to-patient variability also affects the concentration of VOCs in the exhaled breath collected⁸⁰. One of the most common methods to concentrate and store breath samples are sorbent tubes. Sorbent tubes contain various types of solid adsorbent materials to best absorb volatiles of interest and there has been several studies assessing storage and time needed to accurately measure VOCs⁷³.

Exhaled breath analysis using the electronic nose (eNose) technique has been shown to discriminate between various pulmonary pathologies, such as asthma and COPD

from healthy subjects⁸¹. eNose is a sensor based technology in which crystal alignment, secondary to the interaction between breath and VOCs, produces an electrical output that can be measured. A profile, or breath signature is created which can then be interrogated further. However, with eNose, individual VOCs are not clearly identified or referenced and thus the technique only allows identification of more generalised 'patterns' of VOCs.

1.6 Overview of Mass Spectrometry

Mass spectrometry is a type of analysis that can determine which molecules make up a sample based on the spectrum created by the ions that are present^{82,83}. Current uses of clinical mass spectrometry include: diagnosis of metabolism deficiencies, particularly when trying to identify inborn errors of metabolism⁸⁴, measurement of biomarkers, enzymes or hormone levels⁸⁵, and for toxicology testing⁸⁶ in body fluids.

Highly sensitive mass spectrometry techniques, such as Proton Transfer Reaction-Time of Flight-Mass spectrometry (PTR-TOF-MS), selective ion flow transfer (SIFT-MS) and Gas Chromatography-Mass Spectrometry (GC-MS) have been developed to measure VOCs and can potentially be used to more accurately differentiate between disease processes⁸⁷. These mass spectrometry-based techniques have increased the capability to detect even the smallest of differences in compounds and could potentially be harnessed for use in clinical diagnosis.

1.7 Mass Spectrometry techniques

1.7.1 Gas Chromatography Mass Spectrometry

In 1952, James and Martin conceptualised GC-MS⁸². Components of a gaseous mixture are separated in time and the compound structural information and identification is obtained. A sample in a gaseous state is transferred from the injector site (heated inlet port) via the use of inert carrier gasses, such as helium and nitrogen, through the column into the detector⁸⁸. The basic principle behind GC is that the 'mobile phase' constituents of a mixture of gas can be separated by passage over a solid stationary phase occurring within the column⁸². This separation between the two phases, and differing interactions between each compound in the stationary phase allows for identification through the individual retention times⁸⁹. The column is within an oven at a high stable temperature to allow for the elution process⁹⁰. Those compounds interacting most strongly take longest to reach the detector, leading to the

production of a plot known as a gas chromatogram⁹⁰. The other common instrument for GC-MS analysis is the quadrupole mass detector, attached after the GC column, through an interface⁸⁸. An interface is the device for transporting the effluent from the gas chromatograph to the mass spectrometer⁸⁹. This must be done in such a manner that the analyte neither condenses in the interface nor decomposes before entering the mass spectrometer ion source⁸⁸. A quadrupole mass detector instrument has four precisely machined rods that the ions must pass between to reach the detector⁸⁸. The effluent from the GC is ionised by the electrical source resulting in almost exclusively positive ions (electron ionisation)⁹⁰. Ions enter the quadrupole rods along the z-axis after being drawn out of the ion source by a potential (typically a few volts)⁸⁸. The ions entering the quadrupole are sorted by imposing electric fields on diagonally opposed rods⁹¹. By sweeping the electric field voltages in a fixed ratio, usually from low to high voltages, ions of successively higher masses follow a stable path to the detector⁹⁰. At any given field strength, only ions in a narrow mass-to-charge ratio (m/z) range reach the detector. All others are deflected into the rods⁹⁰. The quadrupole mass detector then measures the m/z of gas phase ions and provides a measure of the abundance of each ionic species⁸⁸. The measurement is calibrated against ions of known m/z . In GC-MS, the charge is almost always 1, so that the calibrated scale is in Daltons or atomic mass units⁸². All mass spectrometers operate by separating gas phase ions in a low pressure environment by the interaction of magnetic or electrical fields on the charged particles⁹¹. The most common mass spectrometers interfaced to gas chromatographs are the quadrupole (described above) and magnetic-sector instruments⁸⁹.

To concentrate the collected samples, several devices are available onto which samples are transferred⁹². Thermal desorption tubes contain absorbent material that can retain and entrap a single sample to be analysed⁹². In thermal desorption tubes, the adsorbents are arranged in order of increasing adsorbent strength, from sample inlet to outlet⁹¹. The largest molecules in the sample are trapped by the first bed of adsorbent⁹³. Smaller molecules are trapped by successive stronger beds. The thermal desorption tube is then thermally desorbed, to transfer the analytes to the focusing trap, before the second desorption stage transfers them to the GC instrument⁹³. Figure 1 below shows the overall schematic as described above and pathway of the VOCs.

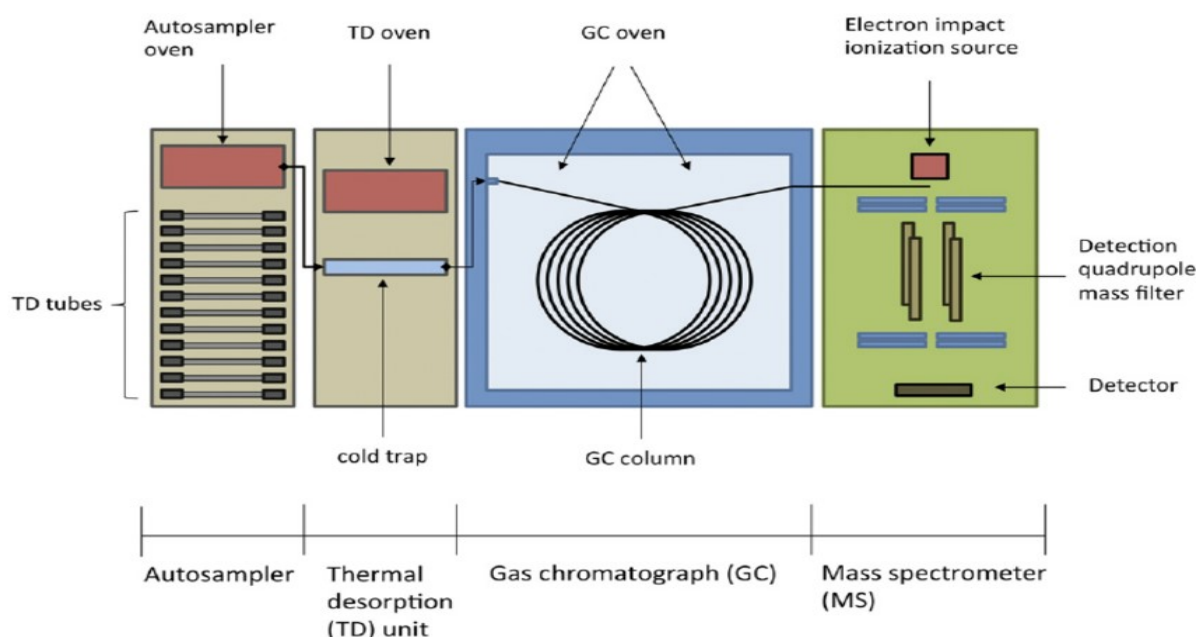


Figure 1: Schematic of the agilent TD GC-MS instrument and pathway of VOCs through the instrument as described by Materic *et al Applications in plants sciences* 2015, 3(12): 1500044.

After a sample is analysed through an instrument, a mass spectrum is produced. A mass spectrum is a graphic representation of the ions observed by the mass spectrometer over a specified range of m/z values⁸². The output is in the form of an x,y plot in which the x-axis is the mass-to-charge scale and the y-axis is the intensity scale⁸². If an ion is observed at an m/z value, a line is drawn representing the response of the detector to that ionic species⁸⁸. The mass spectrum will contain peaks that represent fragment ions as well as the molecular ion. Interpretation of a mass spectrum identifies, confirms, or determines the quantity of a specific compound⁹¹. By running standards of tentatively identified compounds on the same instrument using similar settings, a line should be detected at the same retention time to confirm the compound. Both the intensity and m/z axis are important in interpreting a mass spectrum⁹³. The relative intensity of ions that are observed in a mass spectrum of a pure compound may vary from instrument to instrument⁹¹. In general, most magnetic-sector instruments will produce very similar spectra, provided the ionization conditions are the same⁹³. Quadrupole instruments are tuned to produce mass spectra that are similar to spectra obtained from magnetic-sector instruments. In either case, properly calibrated instruments will have all the same ions at the same m/z values⁸⁸. Thus, two

mass spectra taken over the same mass range using the same ionization conditions should show all the same ions with some variation in the relative intensity of observed ions⁹¹. Extra peaks appearing in a spectrum are caused by impurities or background peaks⁹³.

1.7.2 Limitations of GC-MS

GC-MS is currently used in multiple industries as a common confirmatory test⁹³. Compound separation and identification remains one of the main positives of the technique but it does have some limitations⁸². The size of a spectral peak is proportional to the amount of the substance that reaches the detector in the GC instrument but does not give exact quantification⁹³. Calibration curves need to be developed to allow for semi quantification but developing these presents technical difficulties such as ensuring that a known quantity of compound is transferred onto thermal desorption tube accurately⁹¹. Additionally, no detector responds equally to different compounds so results may be inconsistent⁸². The GC-MS instrument is technically very sensitive and analysis is only reliable with regular maintenance of the instrument⁹⁴.

The choice and maintenance of key components such as the choice of trap and the column will also affect the sensitivity and accuracy towards certain classes of compounds⁹³. The length of the column is crucial for separation but the choice of stationary phase is critical⁹¹. A non-polar substance composing stationary phase is best for separating hydrocarbons whilst a polar stationary phase is essential for oxygenated VOCs such as aldehydes and alcohols⁸². Water from analytes/breath is a major problem with GC work, as it interferes with the column. Therefore, steps to dry and remove water are crucial to reproducible accurate analysis⁸⁸.

1.8 Overview of VOCs in respiratory disease

Emerging advances in mass spectrometry analytical techniques has allowed the potential for measurement of exhaled breath VOCS to help diagnose certain respiratory diseases⁹⁵. Pulmonary TB has been shown to be identifiable with certain volatile markers in breath which are comparable to VOCs collected from headspace of *M.tuberculosis* cultured *in vitro*⁹⁶. *Pseudomonas aeruginosa* infections in cystic fibrosis patients may be identifiable with hydrogen cyanide⁹⁷ and *Aspergillus Fumigatus* with isoamyl alcohol, cyclohexanone, 3- octanone and phenethylalcohol⁹⁸. Exhaled breath

has also been used to identify diagnostic VOCs in non-infective respiratory conditions. The use of sensor technology (eNose) as well as mass spectrometry instruments can allow differentiation of conditions such as asthma⁹⁹, COPD¹⁰⁰, mesothelioma¹⁰¹ and lung cancer¹⁰² from health and also distinguish between different conditions (e.g. asthma *versus* COPD)⁸¹. Fractionated exhaled nitric oxide (FeNO) is a non-invasive marker of inflammation in asthmatic patients. FeNO can be used to monitor airway inflammation and has been included in the British Thoracic Society guidelines as marker to help with initial diagnosis of asthma¹⁰³. This accepted clinical use of FeNO levels indicates that VOCs could be utilised in a range of other clinical decision making processes. Most of the recent literature has shown that a group of VOCs or profile rather than individual markers are typically required to distinguish between clinical groups¹⁰⁴.

1.9 Use of mass spectrometry in bacterial pathogen identification

The application of mass spectrometry (MS) to bacterial identification was firstly proposed by Anhalt and Fenselau (1975), based on the observation that different mass spectra were obtained from bacterial extracts of different species¹⁰⁵. Other techniques such as matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF)-mass spectrometry showed that specific spectral fingerprints could be obtained from whole bacterial cells and have been successfully adapted for the routine identification of microorganisms in clinical microbiology laboratories in the past 10 to 20 years¹⁰⁵.

Conventional microbiological methods for bacterial identification involve quantitative culture, microscopic examination and biochemical testing¹⁰⁶. These procedures can be accurate and reliable, at least for most of the classical microbial pathogens. However, they are also time consuming, laborious, and require specifically trained personnel for correct interpretation of results¹⁰⁶. Conversely, MS is rapid, reliable and easy to perform and has the potential to replace or complement conventional biochemical methods for identification of bacteria, especially for emerging opportunistic pathogens that may be difficult to identify with conventional tests¹⁰⁷. The application of MALDI-TOF-MS tools has been developed in routine clinical diagnosis, including the identification of bacteria at the species, subspecies and strain levels, and the identification of bacterial toxins and antibiotic-resistance type¹⁰⁸. In Chapter 2, a systematic review is presented that summarises the current literature evaluating MS

techniques and VOCs for identification of common community-acquired respiratory infections.

1.10 Use of mass spectrometry in viral pathogen identification

Only a small number of studies assessing VOCs in viral infection have been undertaken. Changes in VOC profiles may result from pathogens themselves, from interactions between host and pathogen or from host immune responses. Limited studies have been undertaken but VOC emissions with Influenza A in swine infection seem to be linked to virus-cell and virus-microbiome interactions rather than to inflammatory host reactions¹⁰⁹. In vitro human rhinovirus (RV) studies by Schivo et al assessing VOC infection of primary human tracheobronchial cells surmised several VOC compounds may serve as biomarkers of HRV infection¹¹⁰. Mechanism of VOC production is unknown. Toll like receptors3 (TLR3) were investigated by Schivo et al and production of VOCs was thought not due to TLR3 stimulation but does require active viral replication and an alternative therapy¹¹⁰. The biochemical transformations accompanying infection result in specific changes, such as altered protein synthesis. Even minor alterations to a cell's genome can lead to differences in its VOC production¹¹¹. VOC induction secondary to apoptotic pathways has been hypothesised¹¹¹ in addition to known oxidative stress¹¹² and lipid peroxidisation.¹¹³ These studies of important respiratory viruses show feasibility in identifying VOC biomarkers for viral infection for point-of-care diagnostics.

1.11 Existing studies that evaluate exhaled breath to determine exacerbation aetiology

There are very few previous studies that have examined the potential for evaluating exhaled breath to determine exacerbation aetiology in COPD. In a single-centre trial conducted in the Netherlands, patients admitted with acute exacerbation of COPD were tested with an electric nose device, eNose(®), and a diagnosis of viral or bacterial infection was obtained by nasal swab PCR or sputum culture respectively. A neural network with leave-10%-out cross-validation was used to assess the data generated. Forty-three patients were included in the study. The eNose was able to distinguish between COPD subjects suffering from a viral infection and those without infection, showing an area under the receiver operator characteristic curve (AUC) of 0.74.

Similarly, for bacterial infections, an AUC of 0.72 was reported¹¹⁴. These preliminary findings provide rationale for the feasibility of breath evaluation to help with early diagnostics in acute respiratory infections. However, an AUC of 0.72-0.74 indicates only a moderately good test and therefore suggests that better strategies require development.

1.12 Aim and Hypotheses

The overall aim of this PhD is to identify VOC profiles that originate from infection with different bacteria and viruses *in vitro* and to investigate whether measurement of these VOC biomarkers in human breath could allow differentiation between bacterial and viral exacerbations of COPD.

1.12.1 Overall Hypothesis:

VOC profiles can distinguish between bacterial and virally induced exacerbations in COPD.

1.12.2 Sub Hypotheses:

- 1) Pathogenic bacteria produce distinct species-specific VOCs that can be measured in the headspace following culture on agar medium.
- 2) Infection of airway epithelial cell cultures *in vitro* with pathogenic bacteria and respiratory viruses will lead to production of virus- and bacteria-specific VOCs that can be detected in the headspace
- 3) Experimental rhinovirus challenge in healthy subjects with subsequent breath sampling will identify similar VOCs to those identified rhinovirus-stimulated airway epithelial cell cultures.
- 4) Targeted breath analysis in COPD patients during acute exacerbation will allow identification of pathogen-specific VOCs identified in sub-hypotheses 1-3.

2 Review VOCs in common community infections

2.1 Introduction

This chapter is a systematic review of the literature to identify VOCs linked to common community respiratory pathogens. This review has focused on two bacteria and viruses, which account for nearly 80-90% of COPD exacerbations; *S. pneumoniae*, *H. influenzae*, rhinovirus and respiratory syncytial virus.

Bacterial strains have a distinct metabolism, which may result in the production of bacteria specific Volatile Organic Compounds (VOCs). Detection and identification of VOCs using analytical techniques has significant diagnostic potential in medicine^{115,116}. Techniques including Gas Chromatography Mass Spectrometry (GC-MS) and selected ion flow tube mass spectrometry (SIFT-MS) for detection, identification and quantification of VOCs produced by bacteria have been shown to be reproducible (i.e. bacteria such as streptococcus and haemophilus shown to produce specific VOCs *in vitro*)¹¹⁷⁻¹¹⁹.

Measurement of Volatile Organic Compounds in exhaled breath could potentially be used to identify a lung specific biomarker of bacterial infection in COPD. Exhaled biomarkers within the context of respiratory disease have been investigated since the 1970s. Increased concentrations of exhaled NO and alkanes in asthma and COPD have been previously reported¹⁰⁴. In the adult COPD population, recent work by Shafiack *et al* utilised the electronic nose device to analyse breath samples. They were able to distinguish patients suffering from an exacerbation compared to those with stable COPD based on VOC breath profiles¹²⁰. Another study utilising sensory technology was able to discriminate COPD patients colonised with bacteria from non-colonised COPD subjects with Area Under Receiver Operating Characteristics (AUROC curve) greater than >0.8¹²¹. However, these sensor technologies do not critically identify the individual discriminatory metabolites and thus does not provide insights into the role of VOCs in disease pathways. Previous studies in COPD have attempted to use inflammatory markers in blood or sputum to define specific exacerbation phenotypic clusters corresponding with bacterial, viral, mixed and non-infective aetiologies⁵⁴. However, there is considerable overlap between these clusters and many of the elevated parameters are also observed in some patients with stable

disease. Metabolic degradation products including Volatile Organic Compounds (VOC) can be detected in exhaled breath and analysed with highly sensitive mass spectrometry platforms¹²². These compounds have the potential to be used as specific biomarkers of viral and bacterial pathogen presence in the airways.

The aim of this systematic literature review was to identify Volatile Organic Compounds (VOCs) linked to the respiratory infections identified above (as common organisms associated with exacerbation of respiratory conditions) and mass spectrometry analysis.

2.2 Methods

Search strategy:

Bibliographic searches using Pubmed (1946-2018), Embase (1947-2018) and OVID Medline (1946-2018) and Google Scholar electronic data bases were conducted up to and including 11th February 2018 for studies on VOCs and mass spectrometry and infections. Search strategies incorporated several domains of MeSH terms and key words: Volatile Organic Compound/s, infection (bacterial, chest infection, haemophilus, lower respiratory tract infection, pneumococcal, streptococcal, viral), mass spectrometry in multiple combinations and combined by “AND”.

Two reviewers (F Kamal and P Patel) independently screened titles and abstracts of studies identified through the searches. For those fulfilling inclusion criteria, full texts of potentially relevant articles were retrieved and assessed for eligibility in final review. Further potentially relevant articles were identified through the searching of reference lists of included studies. Inclusion criteria were studies employing mass spectrometry in analysis of specific bacteria (*S. pneumoniae* and *H. influenzae*) and viruses (respiratory syncytial virus & Rhinovirus) *in vitro* and *in vivo*. Papers included were those that identified specific VOCs, as well as those that referred to distinguishing by mass to charge (m/z) peaks. Exclusion criteria encompassed all reviews, method papers, letters, comments, perspectives and papers without original experimental data and those that used other non-mass spectrometry analysis platforms such as eNOSE. Both reviewers independently extracted data from the selected studies including primary author; year of publication and mass spectrometry platform and VOCs noted.

Figure 2: *Systematic review overview VOCs and common respiratory pathogens.*

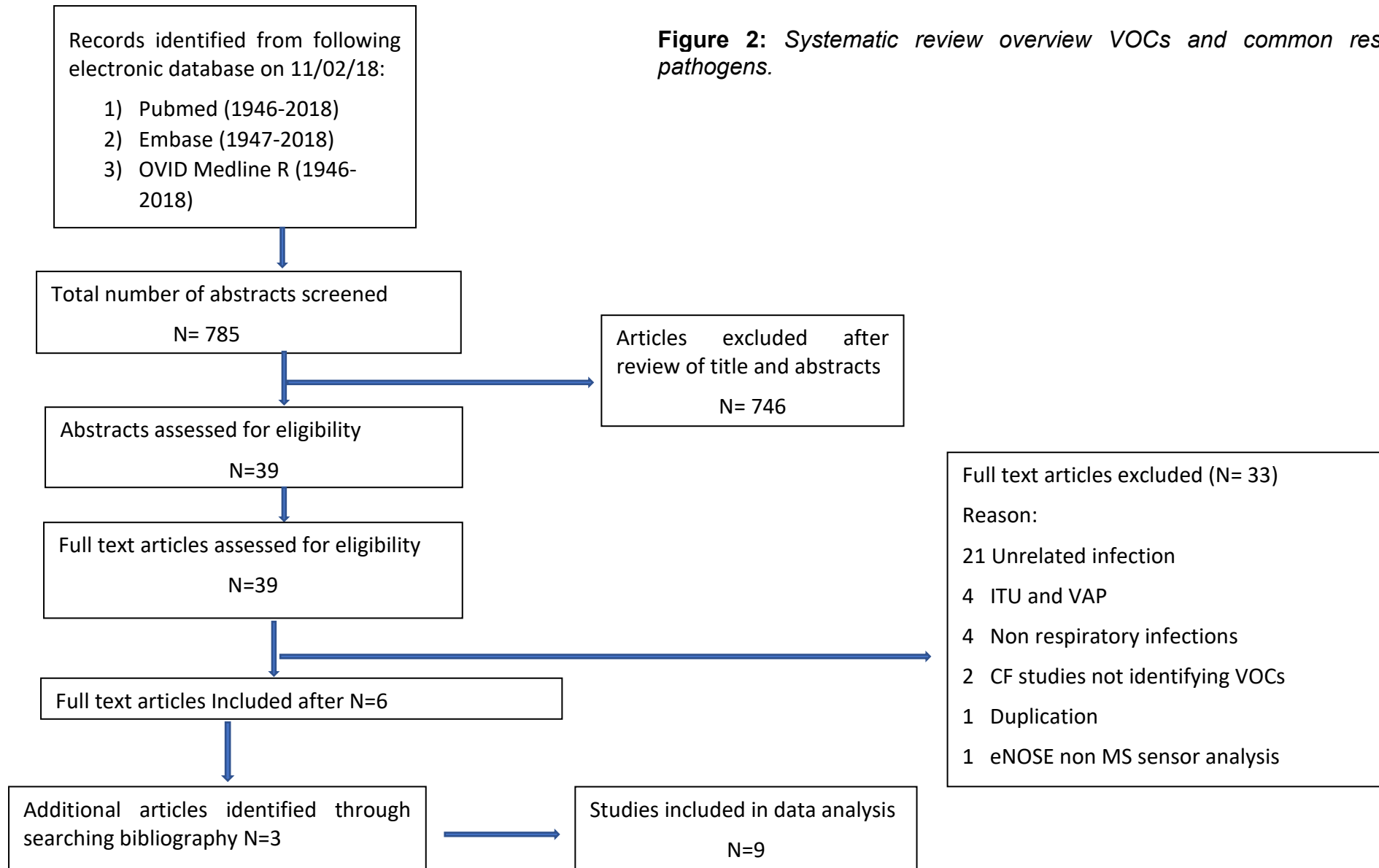


Table 1: A summary of a systematic review of the literature describing VOCs produced from common respiratory infections.

Author and year of Publication	Analytical Platform	Infection(s)	Sample type	VOC established/outcome
Askenov 2014	SPME GCMS	Influenza	<i>In vitro</i>	Alkanes Esters Methylated alkanes Aliphatic alcohols
Chippendale 2014	SIFT-MS (liquid culture medium)	Aspergillus and co infection with <i>pseudomonas aeruginosa</i> (PsA), <i>staph aureus</i> (SA) and <i>streptococcus</i>	<i>In vitro</i>	<i>S. pneumoniae</i> - ammonia, methanol, ethanol, propanol, acetaldehyde, butanal, pentanal, acetone, methanthiol, dimethylsulphide, hydrogen cyanide
El Qader 2014	SPME GCMS	Bacteria(Legionella, morexalla, haemophilus, mycoplasma) Viruses(influenza A/B, Adenovirus, RSV, parainfluenza)	<i>In vitro</i>	1 hexanol and 1 heptadecene increase with viruses
Schivo 2014	SPME GCMS	Rhinovirus infection of transbronchial epithelial cells	<i>In vitro</i>	Differential expression of compounds such as aliphatic alcohols, branched hydrocarbons and dimethyl sulfide
Zhu 2013	SES-IMS	<i>Haemophilus</i> , <i>PsA</i> , <i>Staph Aureus</i> and <i>S. pneumoniae</i>	In vivo Mice	Unique breath prints for infected mice

Author and year of Publication	Analytical Platform	Infection(s)	Sample type	VOC established/outcome
Filipak 2012	TD GCMS (liquid and plate HS analysis)	<i>S. pneuminae</i> <i>H. influenzae</i>	<i>In vitro</i>	<i>S. pneumoniae</i> - butanal, propanol and actaldehyde (34 VOCS) Haemophilus- ester, VSC (28 VOCS)
Junger 2012	IMS and agar plates	Multiple bacteria but also <i>S. pneumoniae</i>	<i>In vitro</i>	3 methyl butanal 3 methyl 1 butanol Benzaldehyde
Alladarce 2006	SIFT-MS	<i>PsA</i> , <i>S. pneumoniae</i> , <i>E.coli</i> , <i>SA</i>	<i>In vitro</i>	Ethanol, formaldehyde, acetaldehyde, indole, dimethyl sulphide, aminoacetophone, hexanal
Scotter 2006	SIFT-MS	<i>S. pneumoniae</i>	<i>In vitro</i>	Acetaldehyde, Ethanol, Acetone, Hydrogen sulphide Dimethyl sulphide, 2aminoacetophonone

2.3 Results of literature search

2.3.1 *Streptococcus pneumoniae*:

In 2006, one of the earliest studies conducted by Allardyce *et al* using SIFT-MS identified the following volatile metabolites from five cultured bacterial strains; ethanol, formaldehyde, acetaldehyde, dimethyl sulphide, trimethylamine, indole, aminoacetophenone and hexanal¹²³. This technique for identification of species using SIFT-MS was corroborated by Scotter *et al*, whom using SIFT, were able to distinguish between *S. pneumonia*, *Pseudomonas aeruginosa*, *E. coli*, *Staphylococcus aureus* and *Neisseria meningitidis*¹²⁴. Comparison of VOC profiles at 24h measuring acetaldehyde, ethanol, acetone, hydrogen sulphide, methanethiol, dimethyldisulphide and 2 aminoacetophenone concentrations (parts per billion) were identified¹²⁴. Further SIFT-MS analysis by Chippendale *et al* evaluating co-infection of *Aspergillus fumigatus* and *S. pneumoniae* identified methanol, propanol, ethanol and acetone as VOCs released when analysing culture headspace¹¹⁸.

Filipak *et al*, using GC-MS, identified 34 VOCs from *S. pneumoniae* bacterial culture headspace analysis¹²⁵. The most significant compounds included; 1) alcohols: methanol, ethanol, 1-butanol, 2) aldehydes: acetaldehyde, propanal, methacrolein, 2-butanal, 2-methylpropanal, 3-methyl-2-butanal, 3) Esters: ethyl acetate, methyl methacrylate 4) hydrocarbons; 1,3butadiene, isobutene, 2 butene, isobutene-2 methyl-1-butene. 5) Ketones: acetone, 2-butanone, 2,3-butanedione, 2-pentanone, 4-heptanone. 6) Volatile sulphur compounds: dimethyl sulfide, carbon disulfide. 7) Acetic acid and furan.

Utilising Ion Mobility Mass Spectrometry (IMS), Junger *et al* demonstrated the ability to differentiate 15 bacterial strains by their respective VOC patterns. VOCs specific to *Streptococcus* included 3-methylbutanal, 3-methyl-1-butanol, benzonitrile and benzyl nitrile¹²⁶. In vivo murine studies of exhaled breath and bronchoalveolar lavage (BAL) were able to differentiate *S. pneumoniae* and *H. influenza* from other bacterial infection using Secondary Electrospray Ionization- Mass Spectrometry. Three distinguishing m/z peaks (m/z = 46. 59 and 74) unique to *S. pneumoniae* were identified¹²⁷.

2.3.2 Haemophilus influenzae

Filipak *et al* identified 28 VOCs linked to *H. influenzae*. There was significant overlap of the same VOCs related to *S. pneumoniae* including 1 butanol, acetaldehyde, propanal, ethyl acetate, methyl methacrylate, 2,3-butanedione, 2-Pentanone, methanethiol, dimethyl sulfide, carbon disulfide, dimethyl disulfide, dimethyl trisulfide, acetic acid and furan. Different VOCs produced by *H. influenzae*; methanol, butanal, methyl propionate, isoprene, 2-methyl-2-butene, 3-methyl-1-butene and 4-heptanone¹²⁵.

2.3.3 Influenza and rhinovirus

Schivo *et al* identified time-dependent release of VOCs related to rhinovirus infected airway epithelial cells. Individual cell lines produce distinct VOCs, indicative of cellular processes¹¹⁰. At 12 hours post-infection, acetone, organic esters and several undetermined aliphatic compounds were noted. At 24 hours post-infection, 2 unidentified hydrocarbons found. At 48 hours, multiple VOCs including dimethyl sulfide, acetic acid, aliphatic hydrocarbons and alcohols, phenol, 2-methyl-5-(1-methylethenyl)-2-cyclohexen-1-ol and 3-phenyl-2-propenal were identified at increased concentrations in the virus cohort.

Askenov *et al* investigated VOC production with β -lymphoblastoid cells infected with 3 live influenza virus subtypes. The patterns of VOC production were unique for each subtype allowing for differentiation. VOCs tentatively assigned to influenza infection included 2-methoxyethanol, thiirane, propanoic acid, butanoic acid (methyl-methyl ester) hexan-3-one (5 methyl) heptan-3-one and octan-2-one.

Most recently, 1-hexanol and 1-heptadecene were found in headspace analysis of viral cultures allowing for differentiation from bacterial headspace¹²⁸.

El Qader *et al* examined the VOCs released from cultures of five viruses (influenza A, influenza B, adenovirus, respiratory syncytial virus (RSV) and parainfluenza 1 virus), three bacteria (*Moraxella catarrhalis*, *H. influenzae* and *Legionella pneumophila*) and *Mycoplasma pneumoniae* isolated colonies. Heptane and methylcyclohexane were identified as significant VOCs related to infectious bacterial activity. 1-hexanol and 1-heptadecene were identified as VOCs that are significantly increased in viral culture compared to bacterial samples¹²⁸.

2.4 Discussion

The systematic review has demonstrated the potential to use VOCs to differentiate between respiratory infections. The majority of studies undertook tentative identification of compounds with limited quantification and no investigation of human samples has previously been carried out. Analysis of bacterial and viral headspace has been shown to be feasible and reliable through several mass spectrometry techniques¹²⁹. Limited in vivo experimentation has shown the utility of breath collection and subsequent VOC analysis in small animal models to identify common community-acquired respiratory pathogens¹²⁷. There remains significant potential to utilise VOCs for point of care breath test analysis for the identification of pathogenic organisms allowing for more appropriate management of patients. Volatile Organic Compounds (VOCs) are produced by virtually all metabolic processes of the body. As such, they have potential to serve as non-invasive metabolic biomarkers. Through the review the study design to be conducted was conceived. VOC qualification and identification was sought and GC-MS was shown to be gold standard in untargeted identification¹³⁰. Although some VOCs had been labelled potentially in keeping with certain infections, the lack of reproducibility meant untargeted approach was planned so not to miss any potential biomarkers as well confirming potential markers through reproducibility. Use of thermal desorption (TD) tubes for various sampling collection and preconcentrations has been shown to be a recognised standardised and reproducible method of VOC collection⁹² and the ability to transfer and transport offers opportunity to collect and analyse numerous sample types (breath, head space) from various sites (collections not limited by location of GC-MS).

A key issue with volatile biomarkers identified to date is the relative absence of inter-laboratory reproducibility. Part of these differences is likely due to the current lack of standardized collection techniques, devices, and data-analysis. Instrumental variability and differences in calibration techniques are also likely to be a source of variance. As this is an emerging field, VOC compound identification continues to be challenging as established repositories are continuously updated. In addition, the concentrations of certain VOCs are present at sub parts-per-billion leading to potential measurement errors with highly sensitive instruments.

The ultimate aim is to provide the scientific evidence for a point-of-care breath test to differentiate viral from bacterial infection and thus guide appropriate antibiotic therapy.

The systematic review demonstrates that specific VOCs are linked to common bacterial and viral pathogens. All the studies utilised only a single mass spectrometry technique and the majority of the studies analysed VOCs from only 'in vitro' samples. There was also limited reporting on calibration and quality control measures in each of these published studies.

The systematic review identified how in isolation invitro work provided the feasibility for a diagnostic test. It helped drive the study design which sought to confirm markers in vitro and to follow through to see their translational applicability in a longitudinal in vivo infection model. The potential benefits for patients are:

- Earlier diagnosis of viral/bacterial infections.
- Better outcomes as a consequence of appropriate treatment.

The potential benefits for the NHS include improvement in appropriate antibiotic use, reduction in healthcare costs through faster diagnosis, and a potential resultant decrease in antimicrobial resistance. Currently, there are limited markers to diagnose COPD exacerbation during the acute setting. Exacerbations are not always infective in aetiology and this leads to over-treatment with antibiotics in both primary and secondary care. Diagnosis is very much driven by clinical judgment and any aids to clinical decision making would be beneficial both for patients and clinicians.

3 Evaluation of VOC production by bacteria cultured on agar plates

3.1 Objectives

The first main objective was to set up a model for VOC collection from agar plates plated with *H. influenzae* and *S. pneumoniae* to identify VOCs respective to specific bacterial species. The aim was to show feasibility of collecting VOCs from this model and also determine if differentiating between two common pathogenic bacterial species could be achieved using a bespoke VOC collection method and subsequent analysis using untested GC-MS methodology.

3.2 Introduction

Previous studies suggest pathogenic bacteria produce distinct pathogen-specific VOCs that can be measured in the headspace following culture on agar medium. This section provides methodology to explain how VOCs from agar plates plated with *H. influenzae* and *S. pneumoniae* are collected. This would provide the initial data to confirm ability to establish a VOC collection model and provided invaluable experience in GC-MS methodology and analysis which would be developed further when infecting cells with bacteria and viruses in later studies.

3.3 Material & Methods

3.3.1 Plating and culturing bacteria

S. pneumoniae and *H. influenzae* were grown on Chocolate agar plates (ThermoFisher Scientific, UK, LTD) supplemented with Haemoglobin and GCHI (a chemically defined supplement formulated specifically to facilitate the growth of *gonococci*). GCHI enrichment provides the V factor required for growth of fastidious organisms e.g. *S. pneumoniae* and *H. influenzae*. Bacteria were prepared previously within the group by Dr A Telcian and stored in Todd Hewitt (TH) medium preparation (36.4g of TH medium powder, 1000mLs of distilled water, pipette used to stir/transfer placed into a flask which is autoclaved for 115mins at 200°C).

S. pneumoniae and *H. influenzae* were stored as a master stock of 1×10^8 CFU per mL and individual aliquots thawed as required.

Initially *S. pneumoniae* was grown on chocolate agar plates in a Tedlar bags (Sigma-Aldrich Company Ltd, Gillingham, UK) versus not in a Tedlar bags to assess whether encasement had any gross effects on bacterial growth. Plates were inspected and colony growth assessed and compared visually and to see if there was no difference observed between sealing plates within a Tedlar bag or growing as one would normally do when plating bacteria on a plate without Tedlar bag and placing within an incubator.

3.3.2 VOC collection

A VOC method collection subsequently established with all work taken place in a class II microbiological safety cabinet (MSC). Bacteria in serial 10 fold dilutions 10^8 , 10^7 , 10^6 , 10^5 , 10^4 were pipetted onto chocolate agar plates and spread evenly with a plastic sterile spreader. Agar plates were subsequently placed inside 2L Tedlar bags and sealed inside with a thermosealer (two edges of plastic brought together). This allowed for capture and stability of volatiles within Tedlar bags. The bagged plates were placed in an incubator at 37°C for 24h and VOCs subsequently allowed to accumulate within the Tedlar bags with visible expansion of bags noted.

At the designated time point, headspace VOCs which had been collected and trapped within the Tedlar bags for 24h were entrained across a thermal desorption (TD) tube using a SCK hand pump at 50mLs/minute for 3 minutes. This method is similar to that used by group in previous studies¹³¹.

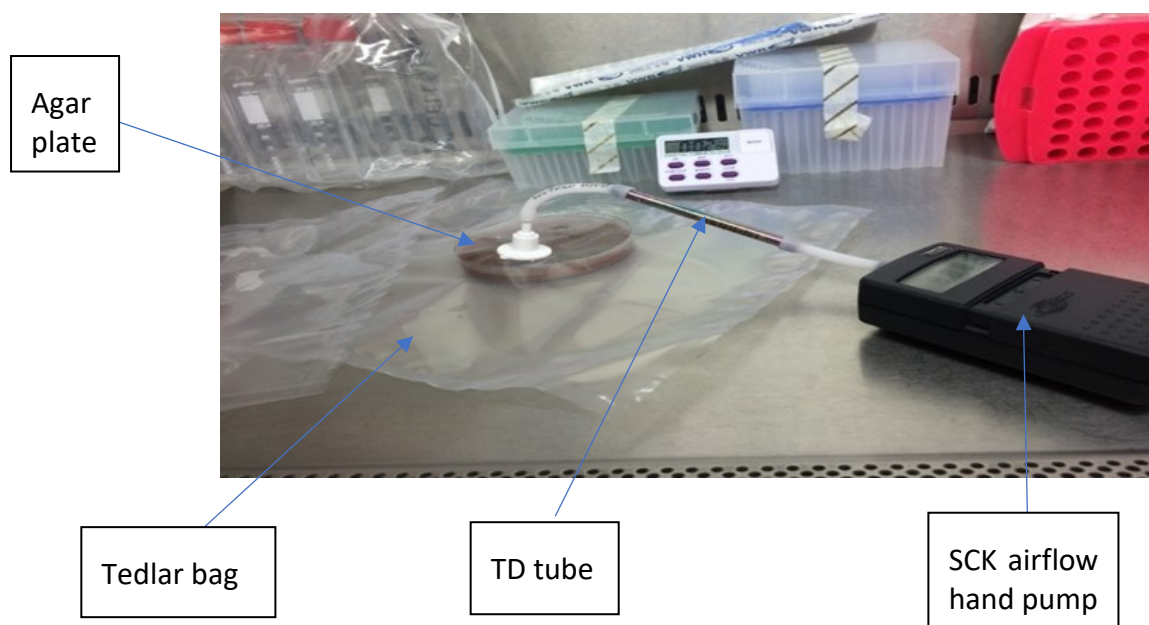


Figure 3: VOC collection from agar plates. VOCs collected from bacteria plated agar plates encased in Tedlar bags. Volatiles transferred onto TD tubes via airflow pump

3.3.3 TD tube preparation

Before headspace of samples were transferred onto TD tubes, some level of preparation was required for the conditioning process. All TD tubes used throughout all studies (pre-packed with 200 mg of Tenax and 100 mg of Carbograph5) were purchased from Markes International (Llantrisant, UK). Prior to sampling, tubes were consistently conditioned at 325 °C for 40 minutes in a stream of nitrogen passed through a hydrocarbon trap (Supelco, US) using a Markes International TC-20 tube conditioner, operated according to manufacturer's recommended conditioning methodology. Once headspace from bacteria samples was transferred (via airflow pump) at the rate of 50mLs/min onto TD tube, they were ran on the Agilent GC-MS instrument.

3.3.4 GC-MS analysis

Instrument settings and the adopted GC-MS methodology were constant for studies undertaken in both the analysis of VOCs from *in vitro* plate and cell infection as well analysis of VOCs from human breath as in other chapters¹³².

Adsorbent tubes) were loaded in the Thermal Desorber's (TD-100, Markes International) sample tray. A single tube was loaded, sealed in the carrier gas flow path and pressurised for a leak testing. Following this, the tube was dry purged for 1 minute

to remove excess moisture and air. VOCs were then desorbed from the tube by a combination of heat (280°C) and nitrogen gas flow (50 mL/min) for 10 minutes and are refocused on the cold-trap (held at 10°C). The cold-trap is then rapidly heated to 290°C with nitrogen carrier gas flow so that VOCs are transferred to the GC column in a narrow band. For separation, the GC column was ramped as follows: initial oven temperature held at 40°C for 4 min, ramped 5°C/min to 120°C with a 1 minute hold, ramped 5°C/min to 200°C with a 1 minute hold, and finally ramped 10°C/min to 240°C with a 4 minute hold. The MS transfer line temperature was 240 °C and electron ion (EI) source conditions were 70 eV at 230 °C. The mass spectrometer was run in electron impact ionisation mode, scanning mass ions 20-250 m/z. An initial 3 minute solvent delay minimised interference from water. The above methodology has been developed and published by previous studies in the group¹³².

3.3.5 Data analysis

Chromatographic data was analysed using the Mass Hunter package (Agilent Technologies) in combination with the National Institute of Standards and Technology (NIST) mass spectral library (2014, Maryland, USA) for compound identification and semi-quantification¹³³.

Data analysis was undertaken using two statistical software R and GraphPad Prism 7. The first step was to perform peak detection and baseline corrections on all analysis output files. Normalization of the calculated peak areas was performed using an area-scaling factor¹³⁴. This rescaling factor used was based on the cumulative area under the detected peaks, since all chromatograms display rather similar profiles this method of normalization is most robust. Next retention times (RTs) of all samples were corrected for chromatographic drifting and lined up. Finally, the output files were merged by combining corresponding compounds based on the degree of similarity of the corresponding mass spectra – by determining the match factor values (MFs) – and similarity of RT. The peaks were then compared between the experimental groups identifying VOCs differentiating between them¹¹⁵.

Untargeted metabolomic studies data are presented in arbitrary units since the studies are untargeted and therefore no calibration standards were analysed. Regarding the detection limits, only peaks with signal to noise ratio higher than 3 were integrated as

per previous literature and recognised GC-MS methodology (this was methodology was repeated for further chapters too)^{130,135}.

3.4 Results

3.4.1 *S. pneumoniae* growth on agar plates

3.4.1.1 Comparison of effects of Tedlar bag encasement on bacterial growth

Initially 100 microlitres containing 1×10^8 CFU/mL *S. pneumoniae* D39 and *H. influenzae* was spread onto chocolate agar plates and incubated for 24 hours at 37°C. Plates were incubated sealed in 2L Tedlar bags with control plates incubated in normal conditions (not encased in a bag). Plates were inspected visually and no difference in bacterial growth was observed with bacterial lawn growth not being affected by being sealed in Tedlar bags.

3.4.1.2 Comparison of headspace VOCs between high dose *S. pneumoniae* versus media and uninfected plate control on chocolate agar at 24h

After having established that encasement of agar plates in Tedlar bags did not have any inhibitory effect on bacterial growth, evaluation of the headspace of *S. pneumoniae* D39 cultured on chocolate agar plates, at an infective doses of 10^8 CFU/mL (n=6) was undertaken.

Through GC-MS analysis, 266 peaks were observed after analysis of the chromatograms for *S. pneumoniae* D39 at 10^8 CFU/mL. Of these 266 peaks, 30 were found to be significantly elevated compared to control media plates. The volatiles were analysed by using statistical package R and Kruskal Wallis analysis was undertaken. Using the National Institute of Standards and Technology (NIST) database and Agilent software, tentative identifications of these peaks have been made, identifying 23 possible compounds.

Standards of several compounds were undertaken to ensure similar retention times and similar mass spectral profiles to allow for identification of volatiles. The 4 most significant VOCs are shown below in Figure 4, several of which were identified in previous studies¹¹⁸.

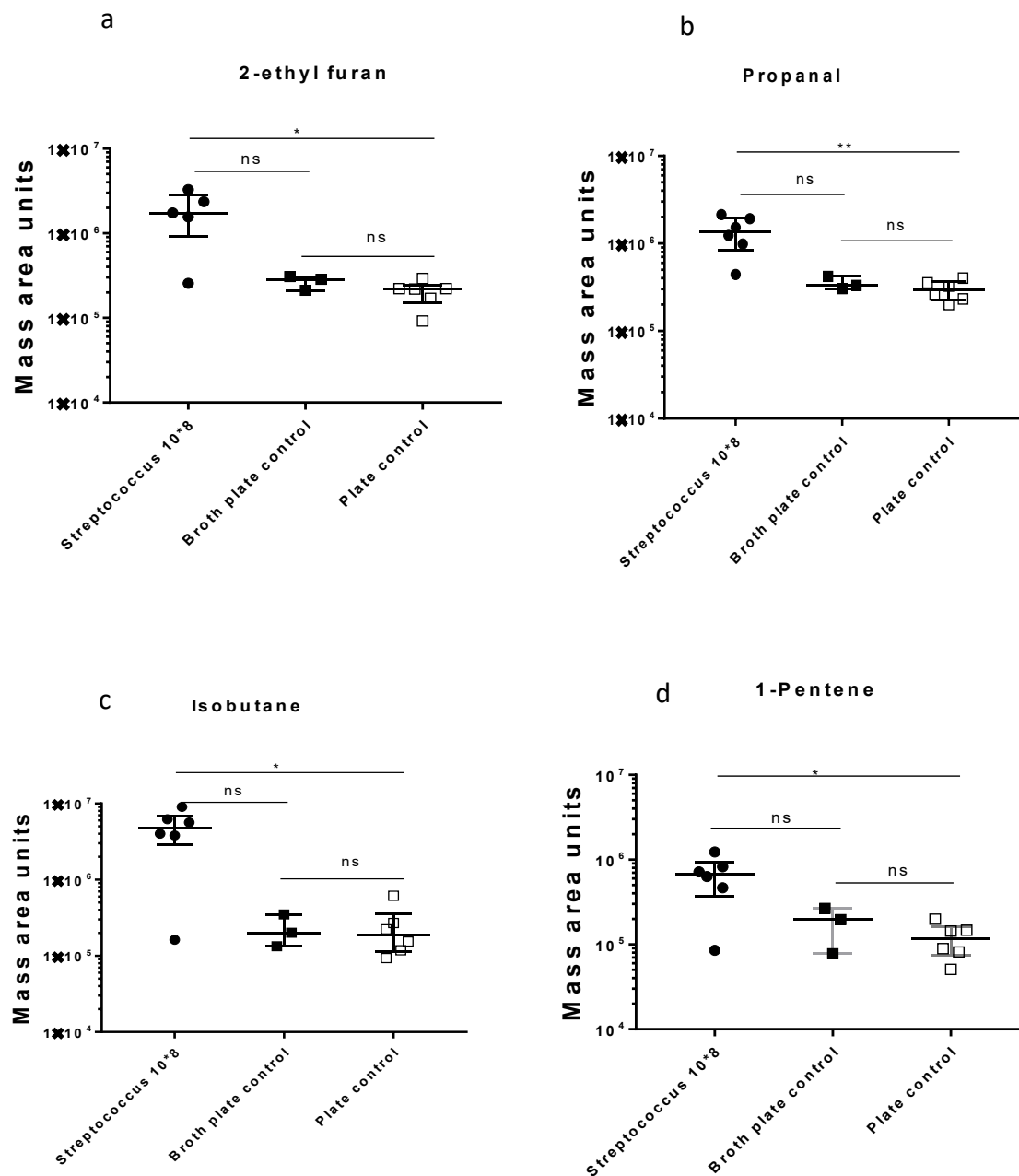


Figure 4: VOC induction by *S. pneumoniae*. 10^8 CFU/mL of *S. pneumoniae* was grown on chocolate agar plates for 24h. VOC induction was measured by GC-MS and compared to agar plates without *S. pneumoniae*. *S. pneumoniae* significantly induced 30 VOCs, 4 are shown above 2-ethyl-furan (a), propanal (b), isobutane (c), 1-Pentene (d) shown). Data was log transformed and analyzed with Kruskal-Wallis with multiple comparisons. Data are presented as median and interquartile range. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ ns=not significant for each dose versus agar plates

The full list of 26 potential VOCs identified include: 2-Butanol, methacrolein, 3-methyl-2-butanol, isobutene, 3-methyl-1-pentene, 1-butanol, 1-pentene, 3-methyl butanal, methyl formate, carbon monoxide, 2-butene, pentane, benzene, propanal, 2 ethyl furan, acetone, 2-methyl 1-propanol, 2-chloro-propane, 3-methyl-pentane, propyne, 2 propenal, 2-propenenitrile and pentanal.

3.4.1.3 Effect of infective dose on VOC production in headspace

The dose dependency of *S. pneumoniae* on VOC induction was next evaluated. *S.pneumoniae* D39 was cultured on chocolate agar plates, as described above at infective doses in a ten-fold dilution series from 10^8 to 10^4 CFU/mL. Eight of the above 26 VOCs were identified to have a dose-dependent increase as the concentration of *S. pneumoniae* bacteria plated increased, there was an increase in mass area under the curve of VOCs produced in 24h. The eight elevated dose dependent VOCs included: Isobutene, Pentanal, 2-butene, propanal, benzene, 2-ethyl- furan, 3-methyl-1- butanol and 1-hexanol. An example of three shown in Figure 5.

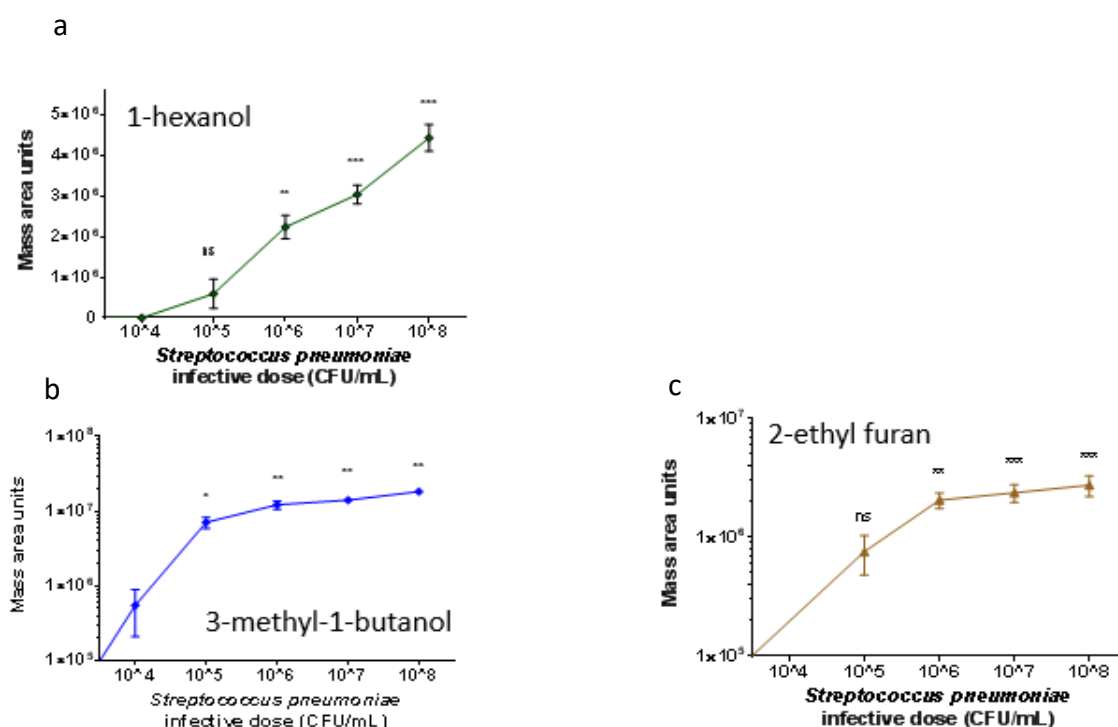


Figure 5: Dose dependent VOC induction by *S. pneumoniae*. Serial 10-fold diluted doses of *S. pneumoniae* were grown on chocolate agar plates for 24h. *S. pneumoniae* gave a dose dependent and significant induction of a) hexanol, b) 3-methyl-1-butanol and c) 2-ethyl-furane. Data shown as mean ($\pm$ S.E.M)* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ for each dose versus agar plates without bacteria as determined by one-way ANOVA with Bonferroni's multiple comparison test.

3.4.2 *H. influenzae* growth on agar plates

H. influenzae was cultured on agar plates (N=6 independent experiments) and headspace VOCs were collected. In total, 262 peaks were identified. Of these, 13 peaks identified through the NIST library cross-referencing were significantly increased, in particular 1,4 pentadiene and Dimethyl sulfide (see Figure 6). Statistical significance was calculated through statistical package R and Kruskal Wallis analysis undertaken to identify significance between the groups.

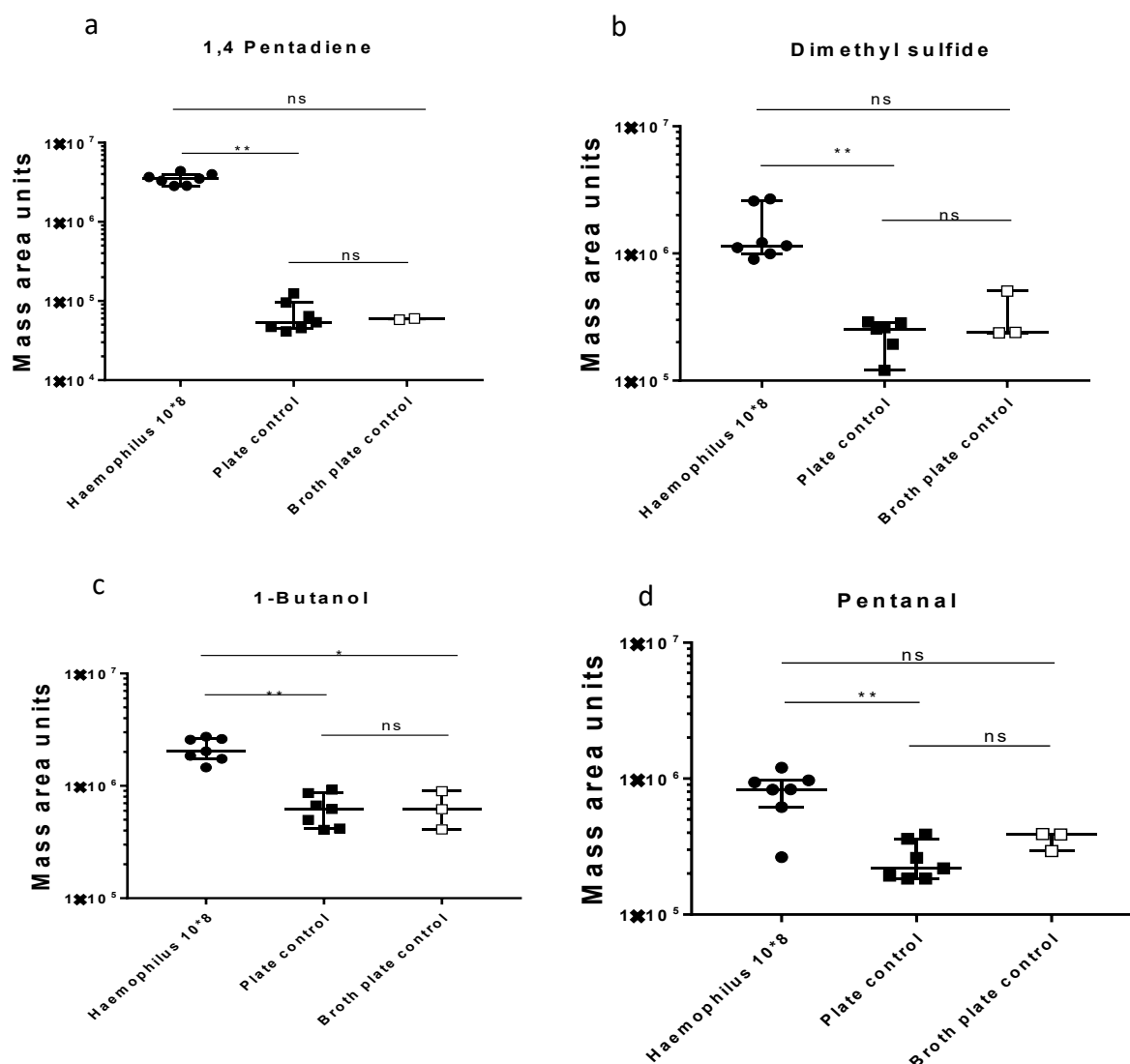


Figure 6: VOC induction by *H. influenzae*. 10⁸ of *H. influenzae* was grown on chocolate agar plates for 24h. VOC induction was measured by GC-MS and compared to agar plates without *h. influenzae*. *H. influenzae* gave significant induction of 13 VOCs (4 shown: 1,4 Pentadiene (a), dimethyl sulphide (b), 1-Butanol (c) and pentanal (d)). Data was log transformed and analyzed with Kruskal-Wallis with multiple comparisons. Data are presented as median and interquartile range. *p<0.05, **p<0.01, ***p<0.001 ns=not significant for each dose versus agar plates

3.4.3 Species specificity and overlap between *S. pneumoniae* and *H. influenzae* VOC production

The following VOCs were induced by both *S. pneumoniae* and *H. influenzae*; 1-Butanol, carbon monoxide, 2-propenal and pentanal. All other VOCs were specific to individual species as seen in Figure 7 below.

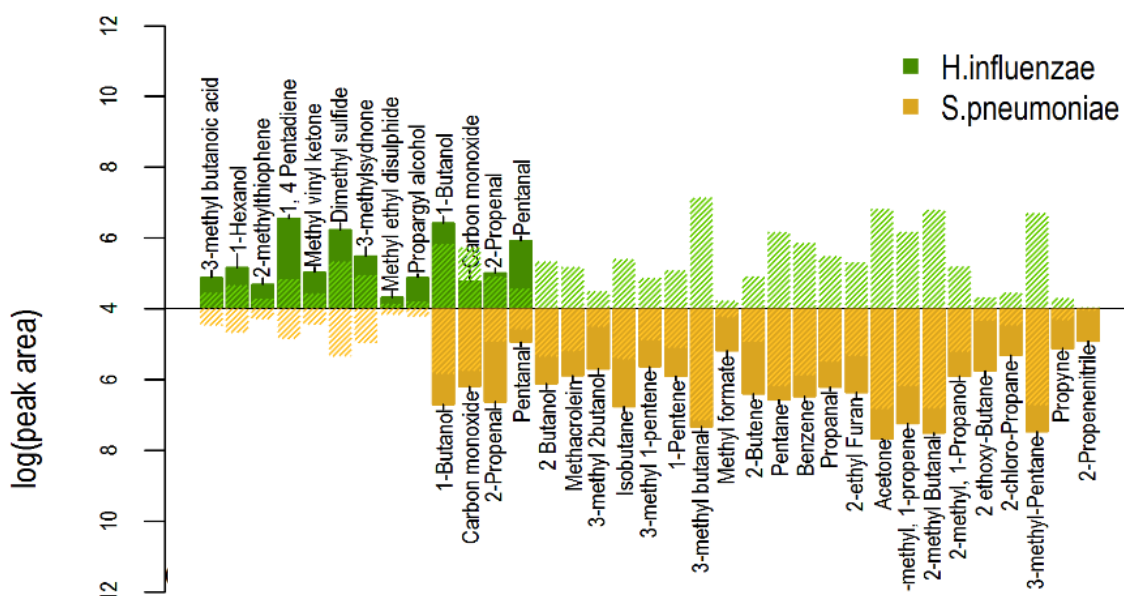


Figure 7: VOC induction by *S. pneumoniae* and *H. influenzae*. 10^8 CFU of *S. pneumoniae* and *H. influenzae* were grown on chocolate agar plates for 24h. VOC induction was measured by GC-MS and compared to agar plates without bacteria. *H. influenzae* gave significant induction of 13 VOCs and *S. pneumoniae* gave 26 VOCs with only 4 overlapping VOCs. The hashed bars are mean values of VOCs found on plate controls. The solid green bars represent mean VOCs released by *H. influenzae* and solid yellow represents *S. pneumoniae* VOC identification.

3.5 Discussion

The current data shows that groups of VOCs such as alcohols, aldehydes, sulphide compounds and fatty acids are increased in the head space of agar plates culturing two common pathogenic respiratory bacteria. As bacteria are grown on agar plates there is no host response and VOCs detected (especially if dose dependent) are therefore produced by the bacteria (or bacteria and media). The previous literature supports the current findings and similar findings have been shown for example by Filiapak *et al*¹¹⁹. Compounds such as alcohols, aldehydes, short chain alkanes and fatty acids can be detected in bacterial infections grown in liquid culture mediums. Other literature has

shown the ability to use mass spectrometry to identify bacteria through differing patterns of VOC induction¹²⁷. In particular, Nizio *et al* using a more powerful analytical separation technique Gas Chromatography-Gas Chromatography time of flight mass spectrometry (GCxGC-TOF-MS) (greater degree of separation and analytical identifications), found similar VOCs in *S. pneumoniae* and *milleri* such as benzene, 3-methyl-butanal, 2,3-pentanedione, alcohols such as 3-methyl-3-heptanol, acetic acid and other fatty acids¹³⁶. Karami *et al* with different bacteria (*Staphylococcus aureus* and *Escherichia coli*) demonstrated that overlying groups of alcohols, fatty acids and aldehydes were detected in identification of bacteria¹³⁷. Alcohols, 1-alcohols are produced through β - or α -oxidation of fatty acid derivatives through acetyl-CoA¹³⁸. Fatty acids are precursors for a variety of important building blocks such as phospholipids, sphingolipids, sterols, as. The degradation of fatty acids yields a high amount of ATP and their detection could be as a result of their presence as an energy substrate facilitating bacterial activity. *S. pneumoniae* and *H. influenzae* induced different VOCs with an overlap of four VOC compounds. Similar groups of compounds as discussed above were elevated in both bacterial groups but different individual compounds were associated with each species. VOC profiles shown in Figure 8 identify four overlapping VOCs but numerous differing VOCs. Bacteria have been shown to produce hydrocarbons including short alkanes, fatty acids and alcohols¹³⁹. These compounds have been thought to provide energy for replication and bacterial growth.

The work in this study is the first time that a dose dependent increase in VOCs related to concentration of bacteria plated has been identified. GC-MS is a semi quantified method of VOC detection (height of peak on GC-MS profile representative of amount of VOC present)⁸². There is a known expectation that different media will produce differing VOCs but the fact that dose-dependent production was observed adds strength to the current conclusion that these VOCs are produced directly by bacterial species. The results from this study have shown similar findings to previously described literature. Table 2 below shows examples of VOCs identified for *S. pneumoniae* and similarities to previous literature; VOCs produced by *H. influenzae* from this study were less comparable to those found previously from the literature. Filipak *et al* identified increase in sulphide compounds such as dimethyl sulfide in *H. influenzae* infection, similar to findings in this study as well as other groups of alcohols, fatty acids and aldehydes¹¹⁹.

Table 2: Comparison of VOCs related to *S. pneumoniae* culture from the present study versus those found previously in the literature.

VOCs found in this study	VOCS found previously in the Literature	Reference
Alcohols 1-butanol 2-methyl-2-butanol 2-methyl-1-propanol 1-hexanol	1-butanol Methanol Ethanol	Junger <i>et al</i> ¹²⁶ Filipak <i>et al</i> ¹¹⁹ Chippendale <i>et al</i> ¹¹⁸
Aldehydes Pentanal Propanal Methacrolein 3-Methylbutanal	Pentanal Propanal Methacrolein 3-Methylbutanal	Chippendale <i>et al</i> ¹¹⁸ Filipak <i>et al</i> ¹¹⁹ Junger <i>et al</i> ¹²⁶
Ketones Acetone	Acetone	Filipak <i>et al</i> ¹¹⁹ Chippendale <i>et al</i> ¹¹⁸
Others 2 ethyl furan Isobutane	Furan 2-methylpropene(isobutene)	Filipak <i>et al</i> ¹¹⁹

The VOCs described in this chapter and chapter 4 need to reviewed with the limitations of the methodology and use of Tedlar bags. Tedlar (PVF, polyvinyl fluoride) is one of the most popular and commonly accepted materials for collecting gaseous samples in general and breath gas samples in particular¹⁴⁰. Studies have shown Tedlar having good recovery (better than 80%) of acetone, isoprene, acetaldehyde and benzene over a 72h period¹⁴¹. However further studies have highlighted the limitations of Tedlar bags. The effect of background environmental VOCs and movement into the Tedlar bags, storage time and associated loss of VOCs beyond 6h, analysis of undefiled Tedlar bags and effects of humidity on VOC recovery have been identified and may subsequently affect VOCs reported¹⁴².

The purpose of this chapter was to assess feasibility with developed methodology to distinguish between two bacterial species, which was shown to be feasible. Certainly, there are limitations of work with failure to assess VOCs produced by commensal bacteria as controls such as *Streptococcus mitis* or *Haemophilus parainfluenzae* and ability to distinguish from them. Commensal bacteria has been identified within the lung with the role of lung microbiota with the role of immunity and health¹⁴³. The utility of differentiating between commensal and pathogenic bacterial species would need to be assessed further in future work to show feasibility of any future clinical assessments in a human population.

3.6 Conclusion

The VOCs collected showed distinct profiles from different bacterial species. Some of the VOCs detected were similar to those previously identified in studies using mass spectrometry techniques within the existing literature. *S. pneumoniae* and *H. influenzae* produced different VOC signatures with minimal overlap. A dose dependent relationship for production of VOCs was observed by *S. pneumoniae*. These data confirm that bacteria cultured in vitro by this method can produce VOCs even in the absence of a host-response.

4 Cell culture, infection and VOC collection

4.1 Objectives

In this chapter, the development of methods for VOC collection from the headspace of airway epithelial cells cultured *in vitro* and infected with the aforementioned bacteria and viruses will be described. This method allowed direct comparison of VOCs infection from bacteria and viruses in an experimental system where host responses would occur.

4.2 Introduction

Chapter 3 has shown the ability to use VOCs as a marker to identify bacterial growth on agar plates. This next chapter examines pathogen specific VOCs in a cell culture system which would recapitulate host pathogen interactions *in vitro*. The hypothesis proposed that infection of airway epithelial cell cultures *in vitro* with pathogenic bacteria and respiratory viruses would lead to production of virus and bacteria specific VOCs that could be detected in headspace. Community pathogenic bacteria *S.pneumoniae* and *H.influenzae* and the viruses RV and RSV were focussed on. These pathogens are the most common pathogens associated with COPD exacerbations¹⁴.

4.3 Methods

Bronchial epithelial cells (BEAS2B cells, an immortalised bronchial epithelial cell line) were prepared and maintained for infection using a number of buffers reagents. These are shown in Table 3 below.

Table 3: Buffers and reagents

Buffer / reagent	Composition
Dulbecco's PBS	Phosphate buffered saline (Sigma-Aldrich)
2% RPMI Medium	500mLs RPMI 5mLs 7.5% Sodium bicarbonate 10mLs Hepes Buffer 10mLs FCS
10% RPMI Medium	500mLs RPMI 5mLs 7.5% Sodium bicarbonate 10mLs Hepes Buffer 50mLs FCS

4.3.1 Cell Culture

BEAS2B airway epithelial cells were maintained in 10% RPMI medium. Cells were cultured at 37°C with 5% CO₂ in 175cm² flasks and split twice per week. For seeding of cells onto plates, 90% confluent BEAS2B cells were washed twice with Phosphate buffered saline (PBS) (5 mLs) and then treated with 5mL trypsin E to dissociate cells. 10mLs 10% FCS RPMI medium was added. Cells were diluted to a concentration of 1.7×10^5 cells per mL and 8mLs added to primaria coated culture dishes (Invitrogen). Cells were infected once 95% confluent.

4.3.2 Rhinovirus infection

RV16 was stored at -80°C as a stock of 1×10^7 TCID₅₀ per mL and thawed when needed. The virus was diluted to a multiplicity of infection (MOI) of 3 in 2% RPMI medium. For each experiment, a batch of RV16 was inactivated by UV irradiation for 30 min or filtered through a 30KDa molecular weight cut-off filter, producing a virus-free filtrate. Infection of the cells was undertaken for 1 hour with shaking at room temperature. Virus was then removed and replaced with 8mLs of 2% RPMI medium. Plates were incubated at 37°C within 2L Tedlar bags for 4, 8, 24 or 48 hours.

4.3.3 RSV infection

For RSV, a similar methodology to that adopted for RV was used. RSV A2 strain at 1 MOI was added in 1.5mLs of 2% RPMI medium per dish and incubated at 37°C, with shaking for 1 hour at room temperature. The RSV was removed and replaced with 8mLs of 2% RPMI medium. Plates were then sealed in Tedlar bags and incubated at 37°C until sampling at 8, 24 and 48 hours.

4.3.4 Bacterial infection

For bacterial infection of BEAS2B cells, *S. pneumoniae* and *H. influenzae* were stored as a master stock of 1×10^8 CFU/mL and thawed when needed. Thawed bacterial stock (200 microlitres) was then diluted in 20mLs of media to give a 1×10^6 CFU/mL solution and subsequently 8mLs of this solution was added to BEAS2B cells on primaria coated culture plates. Plates were then sealed in Tedlar bags and incubated at 37°C with sampling at 8 and 24h.

4.3.5 VOC collection

The VOCs from each of the above set of experiments were collected using a similar technique as described previously in chapter 3. For 3 minutes, the SKC airflow pump entrained headspace contents from within the Tedlar bag at 50mLs/minute onto the TD tube (Figure 8 below).

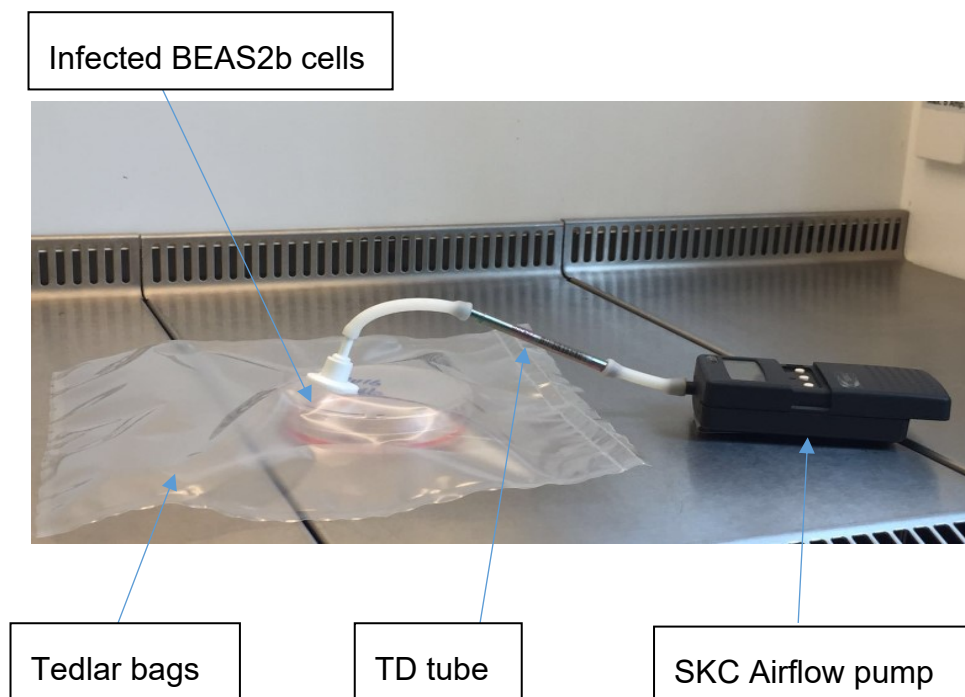


Figure 8: VOC collection from cell model. Infected BEAS2b cells encased in Tedlar bags and VOCs transferred onto TD tubes.

4.3.6 GC-MS analysis

The TD tubes were conditioned as described in chapter 3, and samples were then transferred and analysed using the Agilent GC-MS instrument (Agilent Technologies Ltd, Santa Clara, USA). The agar plate bacteria culture experiments from chapter 3 helped to establish the methodology for the GC-MS settings to allow identification of discriminatory VOCs markers of interest. The TD tubes were desorbed using a Markes TD-100 TD unit (Markes International Ltd, Llantrisant, UK) with a two stage desorption programme, applying a constant flow of helium at 50 mL min⁻¹. In the primary desorption stage, the TD tubes were dry purged for 1 min and heated at 280°C for 10 min. In the secondary desorption stage, the cold trap (U-T12ME-2S, Markes International Ltd, Llantrisant, UK) was rapidly (99°C min⁻¹) heated from 10°C to 290°C. VOCs were transferred from the TD unit to the GC by means of a capillary line heated at 140°C. GC-MS analysis was performed using an Agilent 7890B GC with 5977A MSD (Agilent

Technologies Ltd, Santa Clara, USA) equipped with a ZB-642 capillary column (60 m × 0.25 mm ID × 1.40 µm df; Phenomenex Inc., Torrance, USA) with helium and nitrogen used as the carrier gas (1.0 mL min⁻¹ flow rate). The column temperature settings were the same as those described within chapter 3 subsection 3.3.2. (GC-MS analysis).

4.3.6.1 Standards

To confirm the identity of VOCs by mass hunter and the NIST library, standards were run using to ensure that the identified compounds had the same retention time, with the same instrumental analysis and methodology set up. For liquid standard loading, 0.5 µL of VOC standard mixture at 100 ng µL⁻¹ was injected into individual TD tubes using a GC syringe via a Markes Calibration Standard Loading Rig (CSLR). The GC syringe needle was kept within the loading rig for a short interval of 20 seconds to achieve complete evaporation of the target analyte. Nitrogen flow was set at 50 mL min⁻¹ to aid the retention of the VOC onto the sorbent beds and remove the methanol solvent from the tube. The TD tube with loaded standard was then run on the GC-MS instrument using the same settings as previously described. Chromatograms were compared to identify compounds through retention times and mass spectral analysis. Pattern recognition analyses were then performed on the processed spectral data using MATLAB (The MathWorks Inc., Natick USA).

4.4 Results

4.4.1 Evaluation of effect of Tedlar bag enclosure on BEAS2B cell growth

BEAS2Bs require a temperature of 37°C and 5% CO₂ for optimum growth. Firstly, I assessed whether changing the environment by encasing the cells in a Tedlar bag would affect growth. The effect of sealing BEAS2B cells in Tedlar bags on cell viability was evaluated both qualitatively by visualisation under a light microscope and quantitatively using an MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium colorimetric assay which evaluates cell activity and thus can be used as a surrogate for cell death. No significant difference in cell viability between cells grown enclosed in a

Tedlar bag versus cells grown outside of a bag in normal conditions was observed (Figure 9).

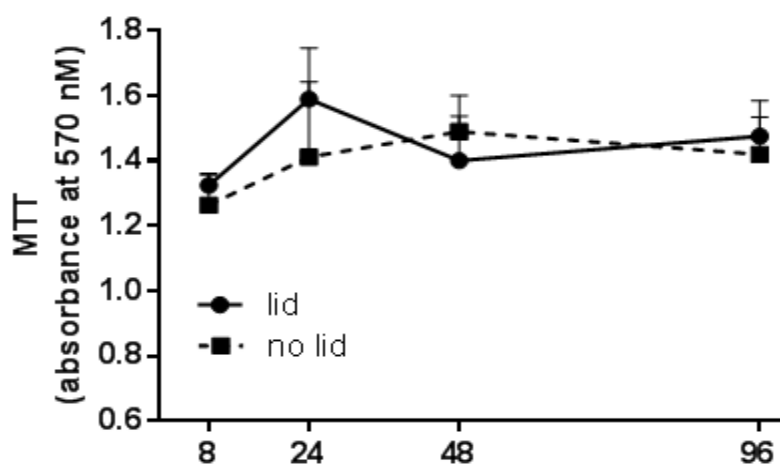


Figure 9: Viability of BEAS2B cells grown in a Tedlar bag. BEAS2B cells in 10mM culture plates were placed in Tedlar bags at 95% confluency (with or without lids) and assessed at different time points. Cells were inspected for viability using light microscopy and an MTT assay.

4.4.2 Comparison of VOC production from BEAS2B cells in medium versus medium alone

To determine if the presence of cells had effects on VOC production, I next determined if VOC production was altered between medium containing BEAS2B cells versus medium alone (without cells). When analysing VOC production from these two groups there was minimal separation (data not shown). This confirmed that the presence of cells has a minimal confounding effect on VOC production.

4.4.3 Infection in BEAS2B airway epithelial cells

4.4.3.1 Cytokine responses

The next stage was to assess production of VOCs in response to RV infection. Initially, to ensure that the stock of RV16 was viable, cytokine responses to RV16 were compared to UV-inactivated RV16, filtered virus or medium. After 24 hours of stimulation, cytokine responses in cell supernatants were determined by ELISA. The inflammatory mediators IL-6, CXCL10/IP-10 and CXCL8/IL-8 were measured as markers of virus-induced inflammation. RV significantly induced IL-6, CXCL10/IP-10

and CXCL8/IL-8 with no significant induction observed for UV-inactivated, filtered virus or medium treated controls.

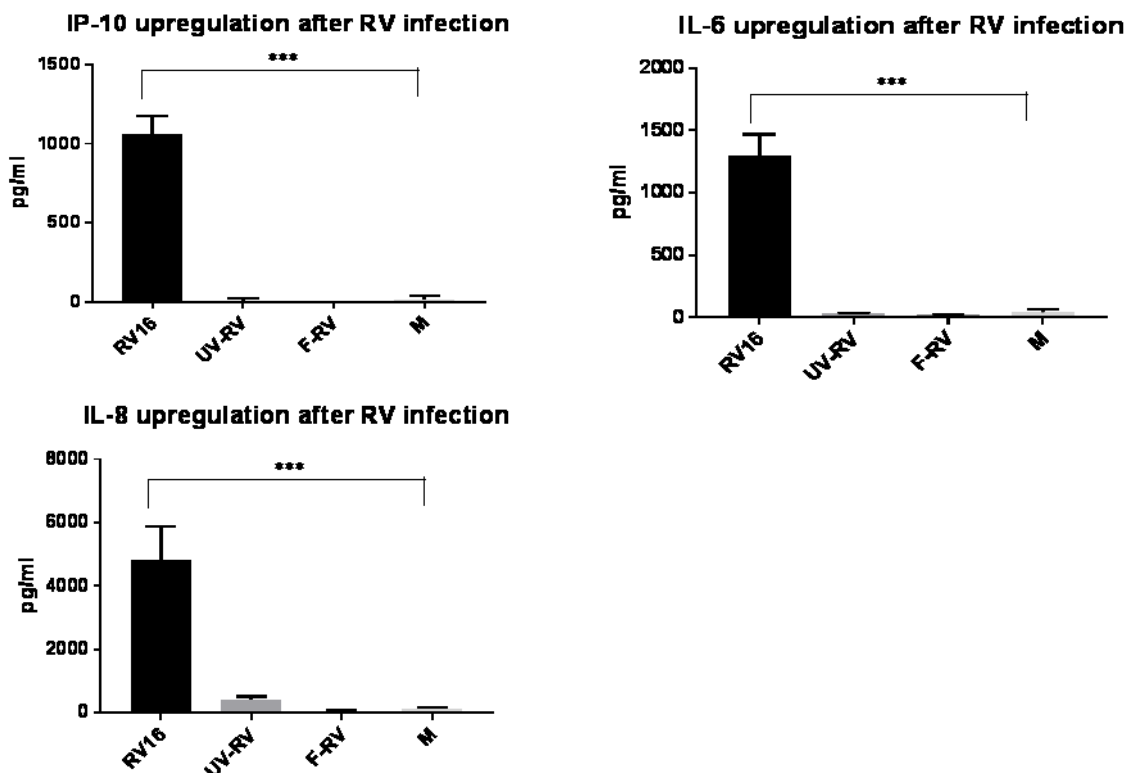


Figure 10: Inflammatory cytokine production in BEAS2B following stimulation with RV-16, UV-RV, filtered virus or medium. Inflammatory cytokines were elevated in supernatant by ELISA following stimulation of BEAS2B cells with RV16, UV-RV16, filtered virus or medium for 24h n=3 independent experiments. Data shown as mean (+/- S.E.M) ***p<0.001 as determined by two-way ANOVA with Bonferroni's multiple comparison test.

4.4.3.2 Measurement of VOCs in headspace of RV infected cells

Having confirmed that RV infection in BEAS2Bs induces cytokine responses, I next proceeded to evaluate VOC production within the headspace of infected cells. Analysis of VOC production at 4 and 8 hours post-stimulation showed no difference between the control and virus infected cells (data not shown). Since all sample measurements were independent (distinct biological replicates), control samples, irrespective of time point of collection, were combined into a single group to increase statistical power.

We performed additional statistical analysis to confirm that the statistically significant VOCs were different due to the infective process and effect of time was negligible

ANOVA analysis of RV-infected versus medium-treated cells identified 14 false discovery rate (FDR) corrected discriminatory metabolites, with good separation. ROC curve analysis indicated an AUC of 0.8 (Figure 11) when comparing RV infected with medium treated cells, indicating a good predictive value.

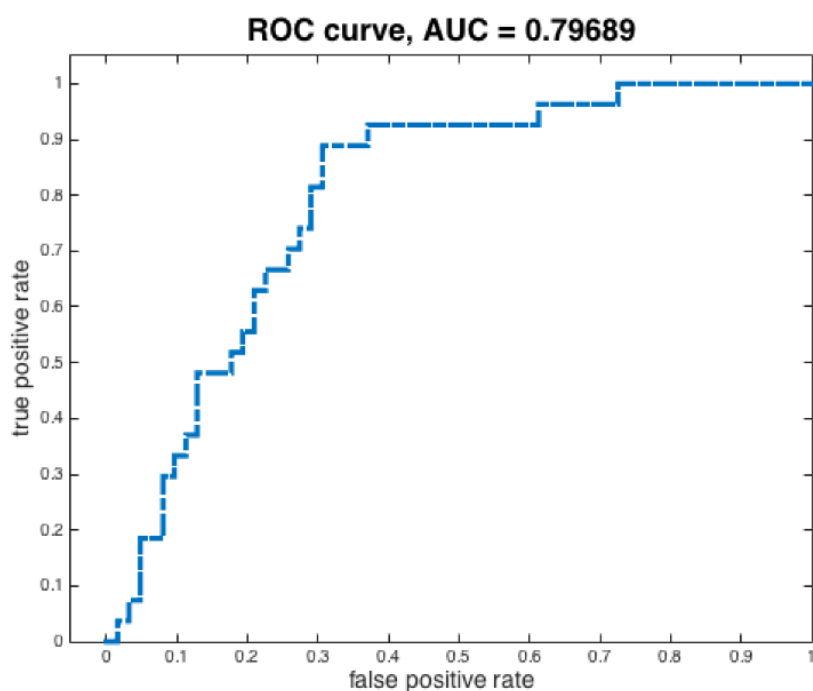


Figure 11: ROC curve showing separation between RV infected cells and control (medium treated cells).

When further interrogating these compounds with the NIST library, only seven of the fourteen provided a good match with the library. These seven compounds were; 2-methyl 2-propanol, cyclobutylamine, acetoin, dimethyl disulphide, decane, 1-chloro-2-methyl-Butane and phenol acetate. The full list of 14 compounds identified is shown in Table 4.

Table 4: 14 discriminatory identified through ANOVA and FDR correction, RV infected versus medium treated cells:

Peak number	P	Q	Log change	RT	%	Compound	M/Z base peak
179	0.000321	0.031275	1.0318	9.27	69%	Poor match (MW 48)	47
197	5.08E-06	0.002312	-	10.07	32%	2 methyl 2 propanol	59
205	1.02E-06	0.000696	-1.8906	10.33	57%	Cyclobutylamine	43
206	4.08E-05	0.009287	-1.0242	10.33	43%	1-chloro-2-methyl-Butane	57
335	0.000252	0.026484	-	14.83	39%	Poor match (MW 119)	33
337	7.35E-05	0.014336	-1.2767	14.93	37%	Poor match (MW 54)	39
460	0.000212	0.026306	0.49521	19.42	100%	dimethyl disulphide	94
463	7.71E-05	0.013148	-2.2581	19.47	8%	Acetoin	45
469	2.34E-07	0.000319	-2.3047	19.8	27%	Poor match (MW 86)	56
657	3.19E-05	0.008707	-	26.69	54%	Poor match (MW 250)	94
718	0.000171	0.023394	-0.8531	28.92	100%	Decane ^a	57
893	1.30E-05	0.004427	-1.1848	33.91	33%	Phenol acetate	94
1220	0.000171	0.025977	1.4178	43.24	76%	Poor match (MW 270)	74
1316	0.000212	0.024114	-1.1673	45.53	56%	Poor match (MW 366)	71

^a indicates identification by matching to NIST MS library and authentic standard

Some of these discriminatory VOCs were upregulated by RV while others were downregulated. Figure 12 is an example of a VOC downregulated by RV infection in BEAS2B cells. Due to poor matching through NIST, it was not possible to identify and label this peak. Graphs (one dimensional scatter plots) have been designed to provide visual assessment of analysis conducted by Mann Whitney U Test of metabolite differences between groups of interest. We showed all data points using one dimensional scatter plots instead of box plots. When using small sample size this is recommended and more informative (able to see distribution of data). Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively. If confidence intervals do not overlap represents $P < 0.05$

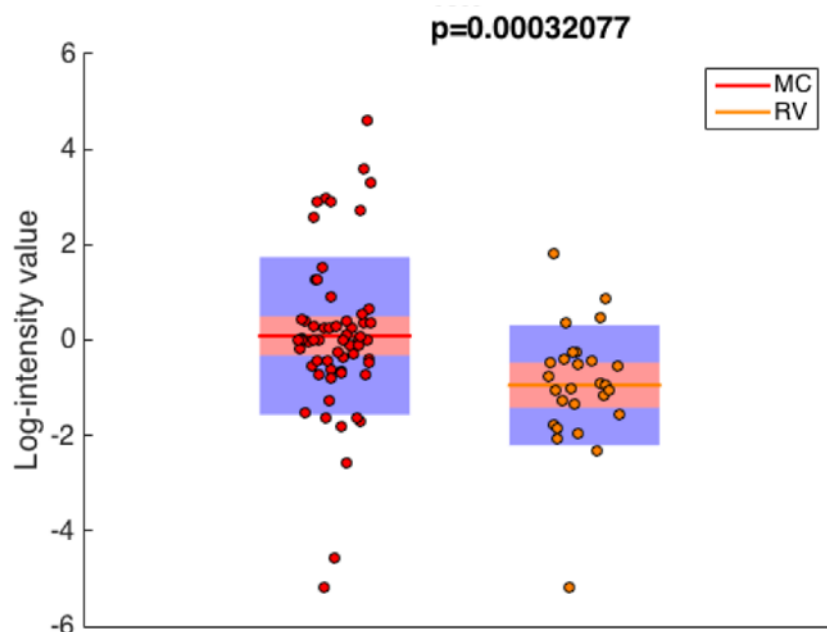


Figure 12: VOC detection of Peak 179 in headspace of rhinovirus infected cells and medium treated BEAS2B cells. MC= medium control; RV = rhinovirus. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively. If confidence intervals do not overlap represents $P<0.05$

Two VOCs identified were found to be upregulated by RV. These were peaks 197 and 718. Peak 718 with a retention time of 29.00min was identified as the compound decane (see Figure 13). Previous studies have shown that decane derivatives can inhibit viral replication, suggesting it could reflect a component of the host-response to RV infection¹⁴⁴.

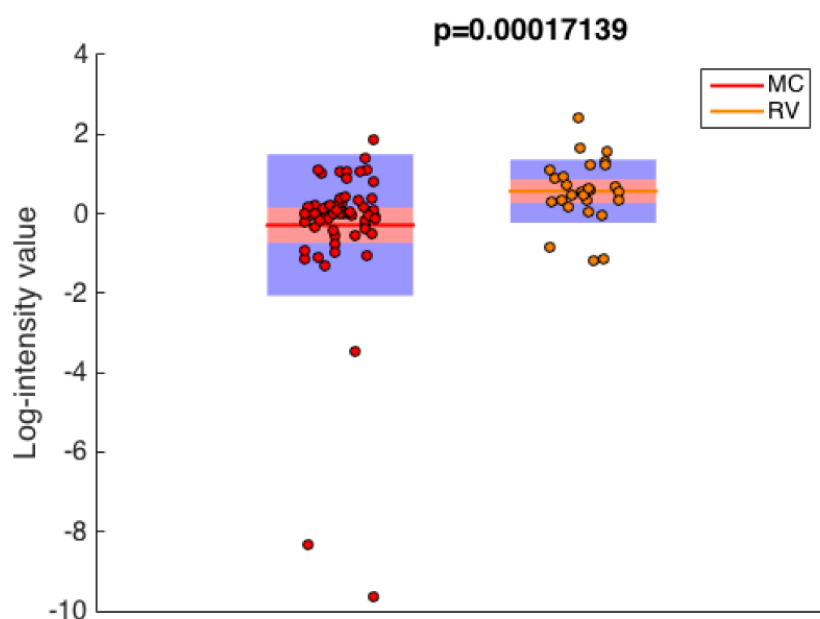


Figure 13: VOC induction of decane in headspace of RV infected cell and medium treated BEAS2B cells. MC: Medium control RV: rhinovirus infected BEAS2bs. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

Peak 197 with a retention time 10.07min was identified as the compound 2 methyl 2 propanol (Figure 14). This compound was also found to be increased in the headspace of bacterial cultures on agar plates (see chapter 3) and may represent a substrate of lipid metabolism (see metabolic pathway outlines in appendix 8.5).

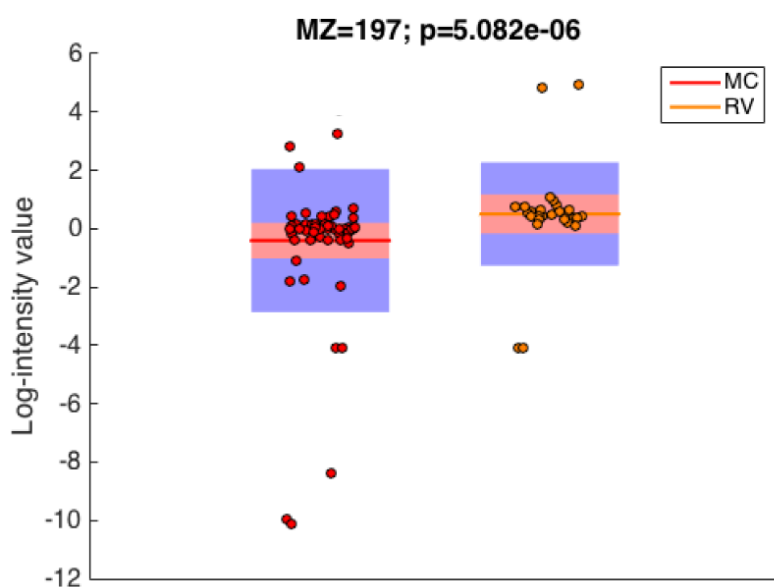


Figure 14: VOC induction of 2 methyl 2 propanol in headspace of RV infected cell and medium treated BEAS2B cells. MC: Medium control RV: rhinovirus infected BEAS2bs. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

Seven of the fourteen identified metabolites showed good matching when reviewed with the NIST library. Standards were also run for the compound decane, which showed a retention time of 29 mins. This matches peak 719 and therefore confirms with certainty that this metabolite is decane (See Figure 15).

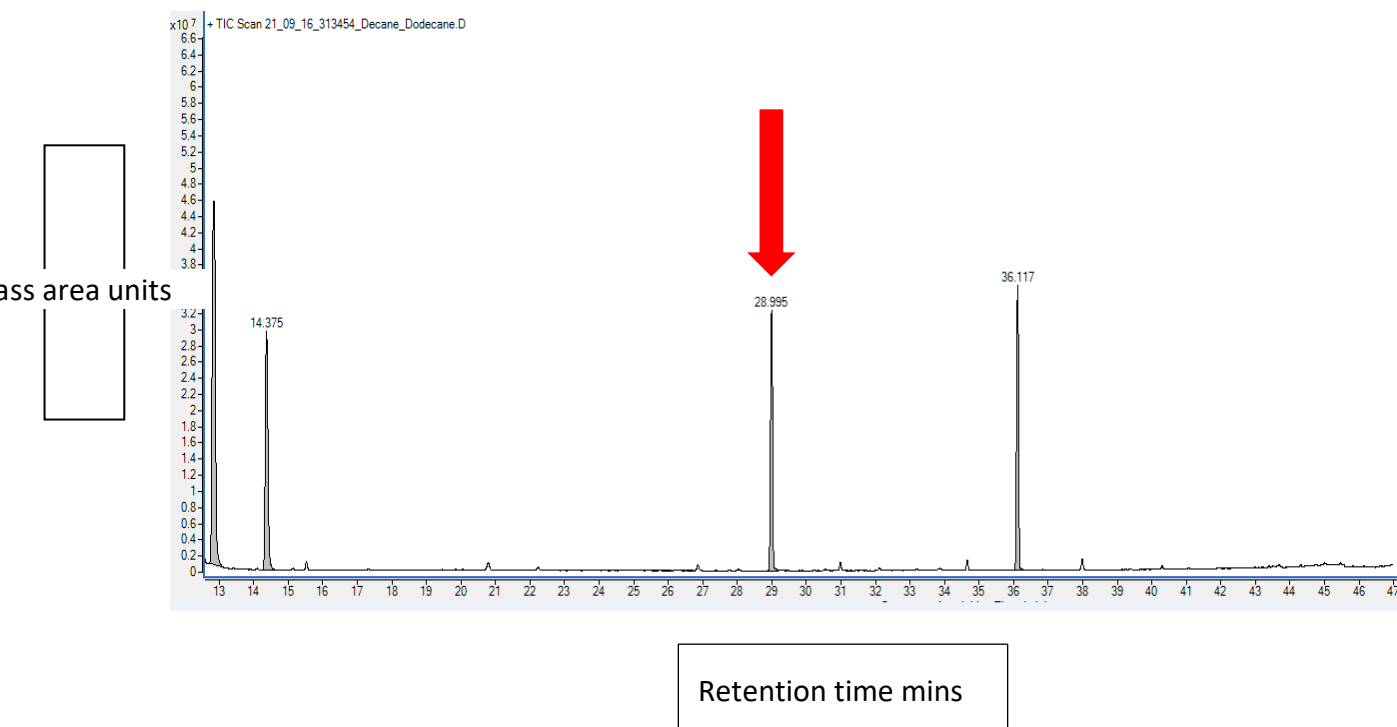


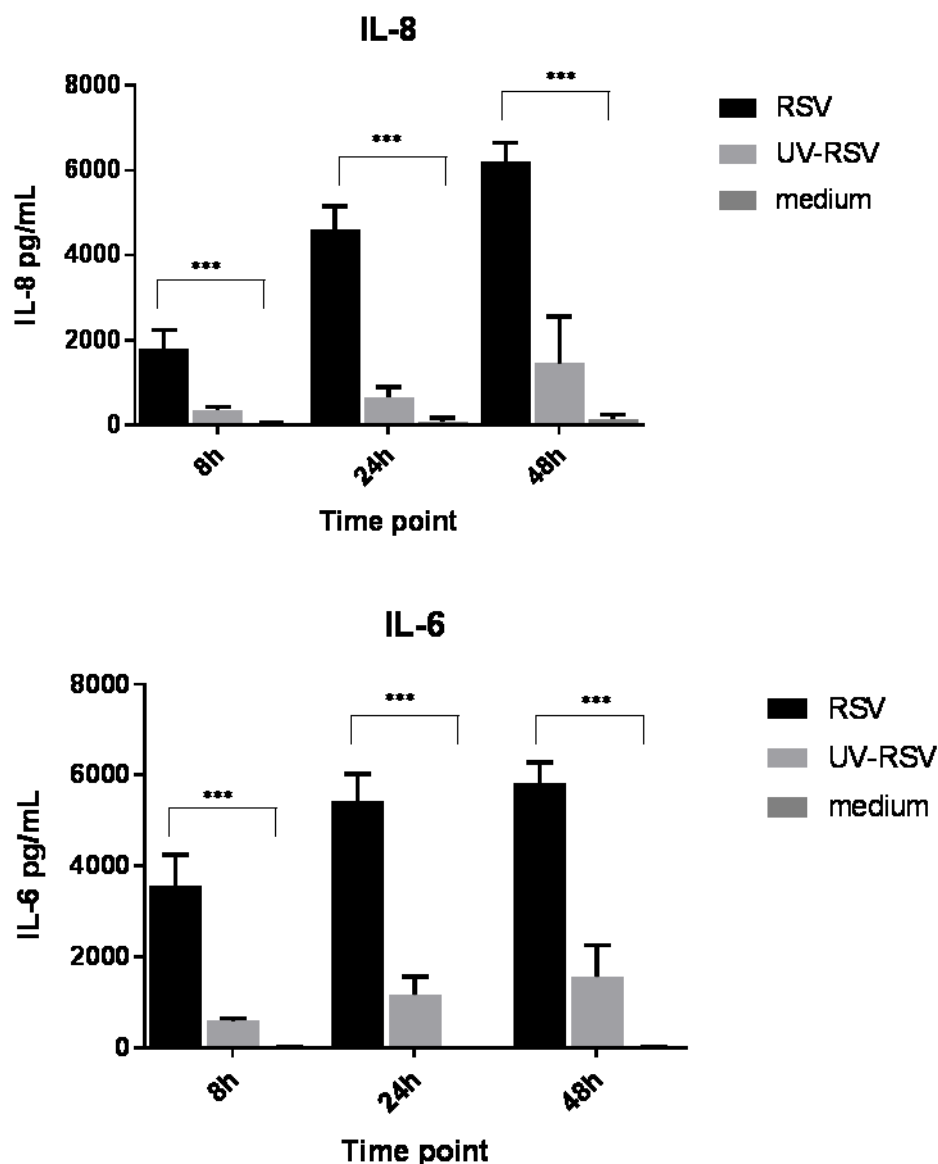
Figure 15: Chromatogram generated by GC-MS analysis of TD tube injected with standard of decane. Time on x axis with spike at 29.00 mins represents decane peak (red arrow)

When comparing RV versus UV-RV or medium treated cells, the VOC analysis at 4 and 8 hour time points showed no separation and thus no discriminatory VOCs were able to be reliably distinguished at these time-points. At the 24hour time- point, when comparing UV and RV treated cells, there was a minimal (5%) variability between the two groups, therefore again poor separations was found on cross validated analysis. The greatest differences were noted at 48h.

4.4.4 RSV infection in BEAS2B cells

4.4.4.1 Cytokine responses

RSV was next studied, another respiratory virus associated with exacerbations of COPD²⁹. Confirmation that the virus batch used induced inflammatory cytokine production by BEAS2B cells was again undertaken by measuring IL-6, CXCL8/IL-8 and CXCL10/IP-10 in cell supernatants at 24 hours post-stimulation. RSV infection significantly induced all three of these cytokines in comparison to UV-inactivated RSV and medium control (Figure 16).



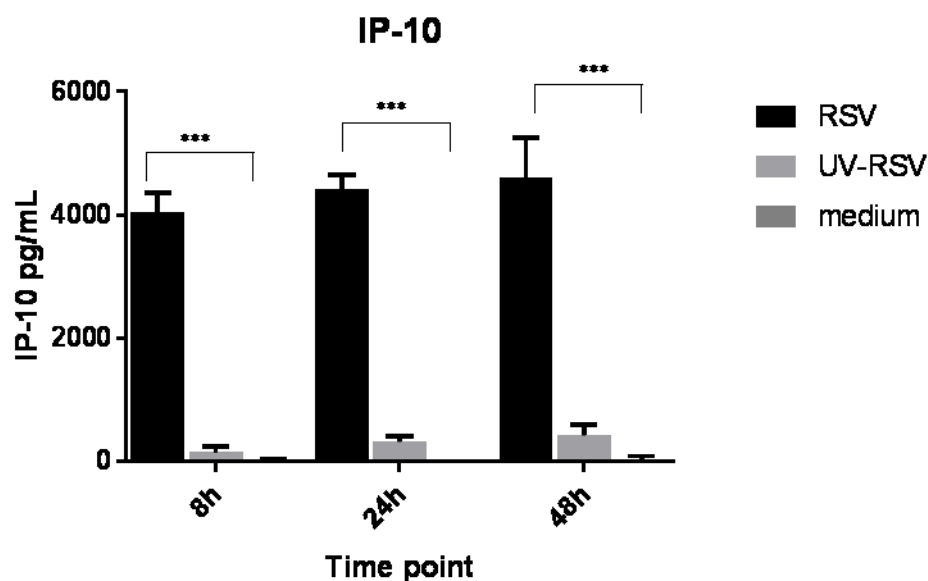


Figure 16: Inflammatory cytokine upregulation in supernatant of RSV stimulated BEAS2B cells. n=3 independent experiments. Data shown as mean (+/- S.E.M) ***p<0.001 as determined by two-way ANOVA with Bonferroni's multiple comparison test.

4.4.4.2 Evaluation of VOC production in headspace of RSV infected cells

A comparison of VOC production from RSV infected and medium treated cells was then undertaken. There was some separation between these two groups with an AUC of 0.76 but no significant differences were observed after correction for FDR. However, when examining the raw data there were around 30 differentially expressed VOCs noted with many showing non-significant trends towards difference between the two groups. The 24-hour time point yielded the most promising results but with variability noted potentially due to batch effects (timing of analysis). To try and interrogate differences further, medium treated cells was used as baseline and compared to all other groups (with baseline adjustment and correction undertaken) to try and improve the statistical power. This allowed the identification of a number of compounds as potential markers of RSV infection (see table 5).

Table 5: 12 discriminatory identified through ANOVA and FDR correction, RSV infected cells versus medium treated control cells:

Peak	P	q	Log Change	Rt	%	VOC	M/z base peak
16	0.018565	1	-2.0672	3.66	71	hydrogen cyanide	28
56	0.033	1	-1.011	5.11	4	Dodeconate	85
210	0.021709	1	-0.77654	10.53	100	diallyl ester oxalic acid	41
275	0.00486	1	-2.5686	12.7	41	acetic acid (acetyloxy)	43
350	0.04804	1	1.2532	15.34	8	Acetic anhydride	43
356	0.014697	1	-1.6249	15.68	56	2-propyl,piperidine	84
358	0.040364	1	1.408	15.68	19	Nitroethane	29
366	0.03795	1	-1.8468	15.77	27	Isobutyl nitrite	43
388	0.009399	1	-0.9268	16.75	4	Poor match (MW 76)	45
401	0.012746	1	-1.7736	17.39	37	3-pentanone	57
836	0.03114	1	1.1592	32.2	100	Phenol	94
1277	0.013915	1	1.6897	44.46	59	Dibutyl phthalate	149

Of these fourteen VOCs, three are known contaminants: Toluene 8, dimethyl-silanol and octamethyl trisiloxane¹⁴⁵ (also found by previous work conducted by the others in the group S.Doran and S. Markez, unpublished observations). The other eleven metabolites showed a very good matching with NIST (forward and reverse matching above 800). A 'differentiagram' (Figure 18) identifies the output from the MATLAB software of those metabolites that, through ANOVA analysis, are shown to be significantly different in RSV infection versus medium treated controls (p value<0.05). In this figure, the colour gradient delineates the degree of significance for easy identification with the X axis relating to peak number. The retention time is found through matching with the master peak list generated to help with VOC identification.

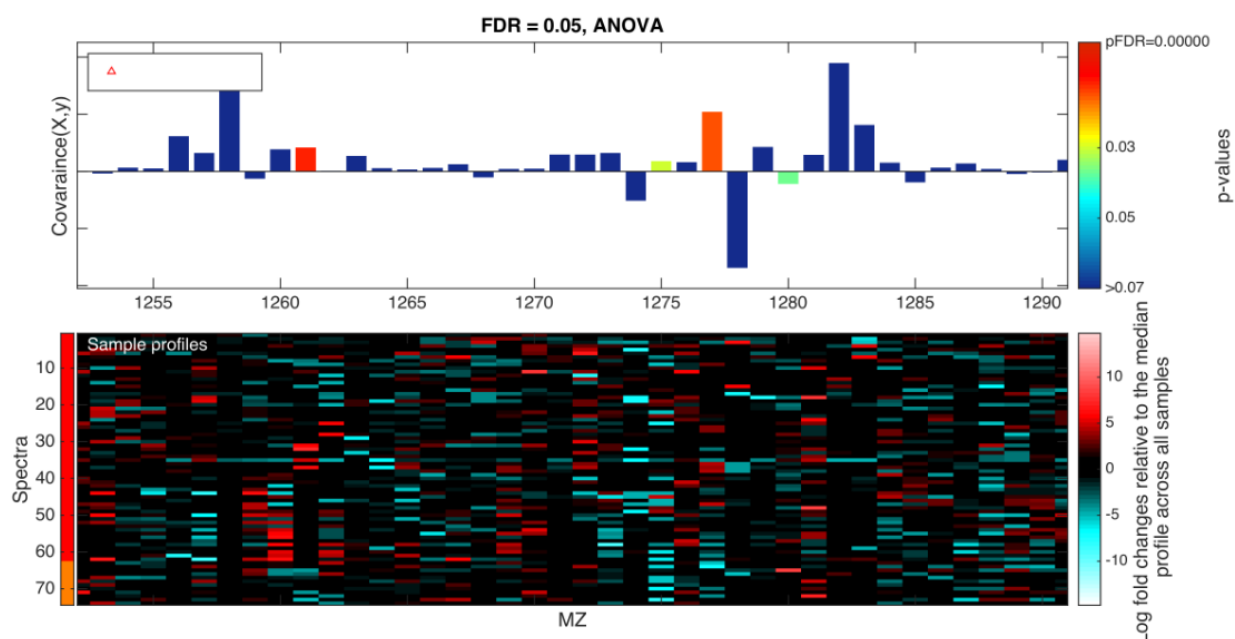


Figure 17: Output from MATLAB software analysis. Example of differentiatogram histogram output. Colour gradient delineates degree of significance from ANOVA analysis undertaken between the multiple samples.

Figure 18 and 19 illustrate two examples of VOCs trending towards increase in RSV infected versus medium treated BEAS2B cell. Notably, RSV upregulated 3 pentanone (Figure 18), a ketone and the fatty acid acetic acid, which has previously been shown to be an energy substrate for replication of viruses¹⁴⁶.

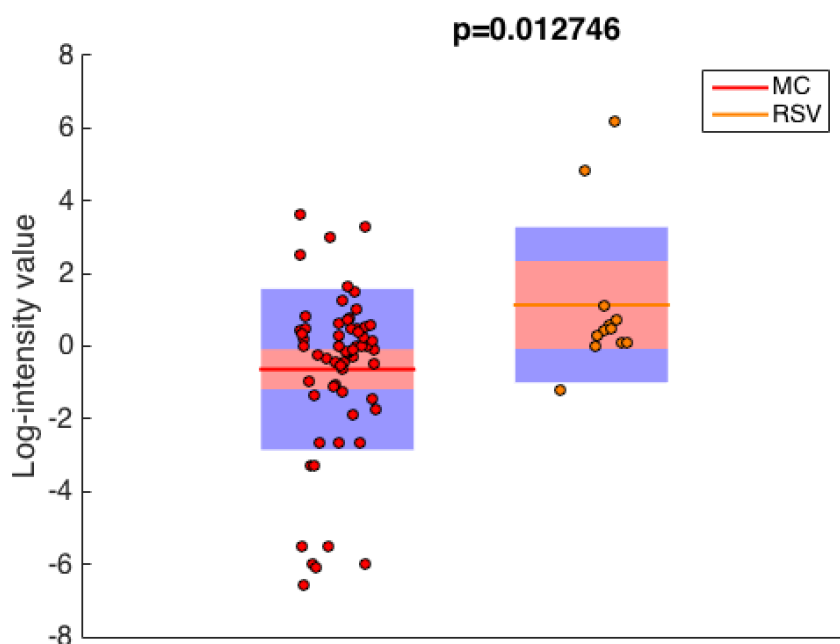


Figure 18: VOC detection of 3-pentanone in headspace of RSV infected cells and medium treated BEAS2B cells. MC= medium control; RSV = RSV infected cells. Data analysed by Mann Whitney U test). Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

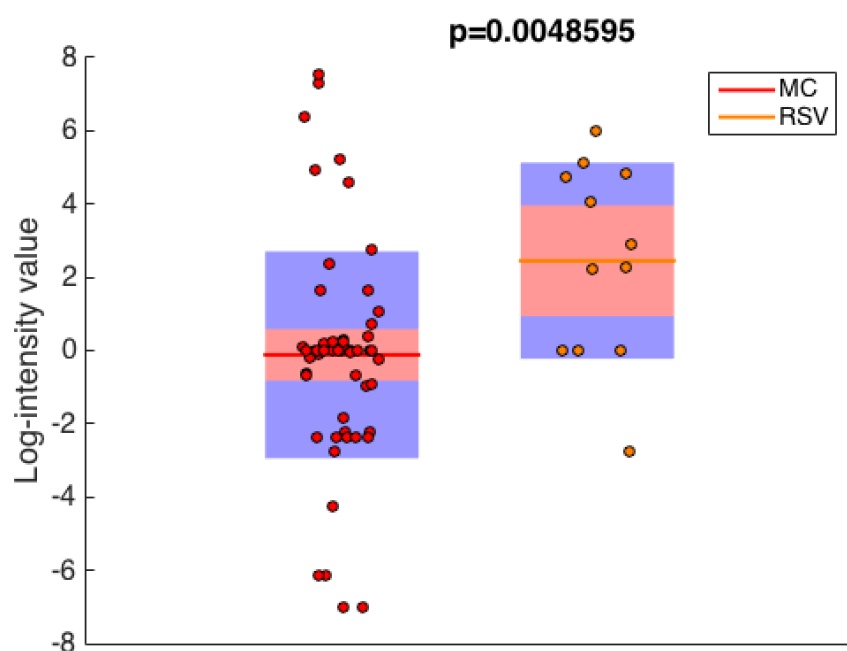


Figure 19: VOC detection of acetic acid in headspace of RSV infected cells and medium treated BEAS2B cells. MC= medium control; RSV = RSV infected cells. Data analysed by Mann Whitney U test). Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

4.4.4.3 Batch adjusted analysis RSV infection

Further time adjusted analysis was undertaken on the same data set focusing on the 24 and 48-hour time points (analysis suggested higher VOC changes at these time points was more likely). The analysis of medium treated cells (control) at 24/48 hours in comparison to 24/48h RSV infected cells gave a 70-75% classification accuracy with relatively good separation noted and Kruskal Wallis analysis undertaken indicated provisional RSV-related VOCs (table 6).

Table 6: Discriminatory VOCs in RSV infected cells versus control medium treated cells). Baseline corrected with time adjustment with Kruskal Wallis analysis and FDR correction; 37 provisional VOC discriminators identified.

Peak	p value	q value	fold change	Retention time	Compound	M/z base peak list
18	0.030902	1	-2.1291	3.75	Unknown	40
39	0.019931	1	-3.1132	4.61	Unknown	113
50	0.008603	1	2.4917	4.92	Unknown	85
67	0.007699	1	-2.7051	5.64	Unknown	51
92	0.020769	1	-2.1488	6.33	Unknown	65
96	0.023742	1	1.4442	6.51	Unknown	45
99	0.035692	1	-2.1866	6.62	Unknown	40
145	0.040421	1	-2.2566	7.99	Unknown	40
153	0.000331	1	-3.3403	8.35	Poor match (MW 74)	59
197	0.035692	1	-0.4133	10.07	2methyl 2propanol	59
301	0.020116	1	-2.5858	13.48	Poor match (MW 230)	81
411	0.012974	1	3.0734	17.85	Poor match (MW 218)	99
463	0.015177	1	-2.1161	19.47	Isopropyl lactate	45
469	0.011684	1	-1.9599	19.80	3,3-dimethyl-oxetane	56
548	0.027423	1	-1.9872	23.05	2-nitropropane 2-chloro	77

566	0.045999	1	-1.8839	23.52	Poor match (MW 98)	43
580	0.035692	1	2.7422	24.00	Poor match (MW 341)	128
589	0.03725	1	-1.4608	24.27	Poor match (MW 127)	43
592	0.038516	1	2.3516	24.31	Poor match (MW 114)	85
665	0.00389	1	4.4606	27.18	Poor match (MW 152)	44
673	0.020769	1	-0.69089	27.50	Poor match (MW 116)	55
743	0.035692	1	0.91063	30.04	Poor match (MW 250)	57
756	0.027423	1	-1.4678	33.95	Poor match (MW 186)	71
873	0.04584	1	0.74991	33.35	Poor match (MW 232)	107
904	0.030087	1	2.3932	34.05	Poor match (MW 144)	71
917	0.020769	1	-2.0788	34.31	Poor match (MW 190)	57
964	0.030087	1	-1.2283	35.70	Poor match (MW 176)	56
978	0.035692	1	-1.0721	35.98	Poor match (MW 59)	59
1035	0.04584	1	-1.0972	37.49	Poor match (MW 590)	57
1057	0.026182	1	-3.1491	37.99	Decanal	57
1115	0.011657	1	-1.025	39.97	Poor match (MW 169)	82
1180	0.035692	1	3.0743	41.91	Poor match (MW 170)	57
1217	0.035556	1	-1.0419	42.98	Entoxeridine	246
1224	0.027279	1	-2.0953	43.29	Poor match (MW 270)	57
1253	0.020863	1	-2.9762	43.87	Poor match (MW 296)	71
1276	0.048939	1	-0.92545	44.29	Poor match (MW 394)	121
1301	0.027423	1	2.1388	45.01	Poor match (MW 98)	55

Only 5 of the peaks identified showed good matching with NIST with two examples shown in Figures 20 and 21, isopropyl lactate and decanal:

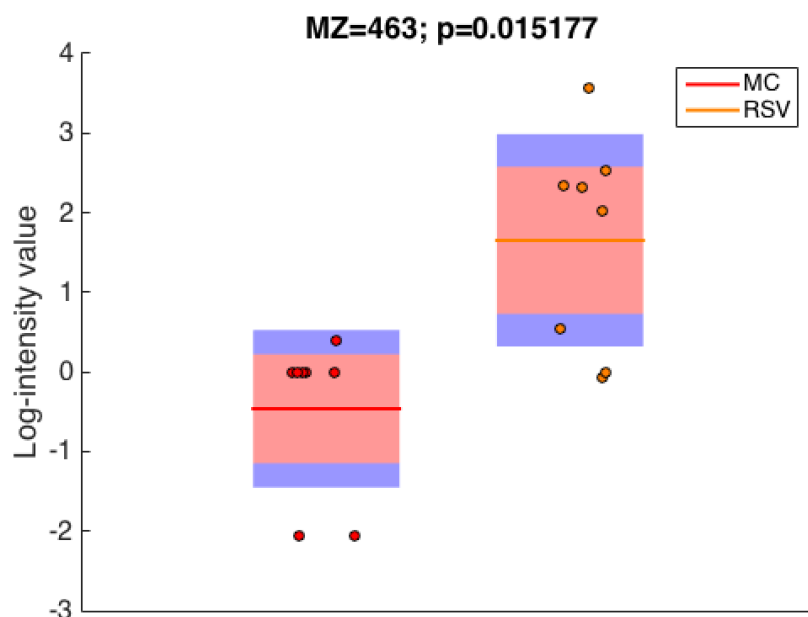


Figure 20: VOC induction of isopropyl lactate in headspace of RSV infected and medium treated BEAS2B cells. MC= medium control; RSV = RSV infected cells. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

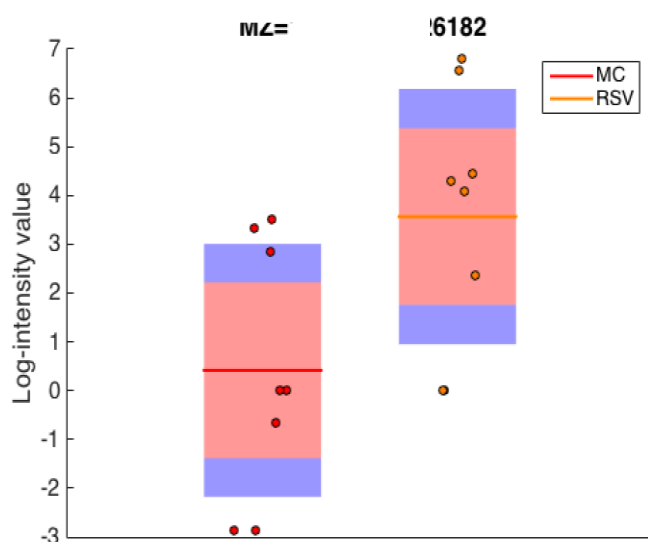


Figure 21: VOC induction of decanal in headspace of RSV and medium treated BEAS2B cells. MC= medium control; RSV = RSV infected cells. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

It is clear from the analysis undertaken that a panel of VOCs is required to distinguish between viral infected versus medium treated cells. There was a significant time-related batch effect and correction was required to help establish discriminative features within the various data sets.

4.4.5 Overlap VOCs between RV and RSV

RV and RSV infected cells were broadly shown to have different VOC profiles. The long chain alkane decane was found to be increased in RV infections compared to shorter chain alkane nitroethane in RSV. There were however some similarities with the

upregulation of alcohols and fatty acids observed for both RV and RSV (e.g. 2 methyl 2 propanol and acetic acid). In addition, nitrogen based compounds such as cyclobutylamine and isobutylnitrile were found to be increased by both RV and RSV infection. In both sets of virus experiments increased levels of phenol were observed (shown in previous vitro VOC studies to be basally released by cells¹⁴⁷).

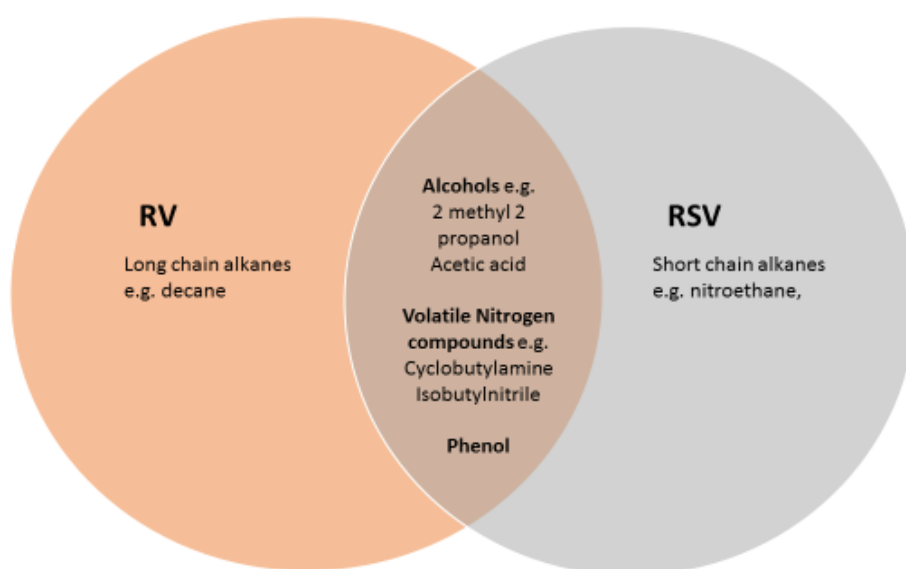


Figure 22: Venn diagram showing the VOC similarities and differences between RV and RSV BEAS2B infection experiments

4.4.6 Measurement of VOCs in headspace of *S. pneumoniae* and *H. influenzae* infected cells

Having studied VOC production in response to virus infection of BEAS2B cells and from bacteria culture on agar plates (chapter 3), I next sought to investigate production of VOCs in response to bacterial infection in BEAS2B cells. Eight VOCs were significantly different between *S. pneumoniae* or *H. influenzae* infected and medium treated cells (see Table 7).

Table 7: Discriminatory VOCs between *S. pneumoniae* and *H. influenzae* infected cells versus medium treated control cells.

Peak	p value	q value	Retention time	Compound	M/z base peak
190	6.38E-06	0.002176	9.82	Poor match (MW 106)	57
197	0.00012	0.018237	10.07	2-methyl-2-Propanol	59
208	3.80E-07	0.000259	10.39	Neopentane	57
263	5.22E-07	0.000238	12.30	3-ethoxy-3-3methyl-2butanone	59
335	2.40E-05	0.004681	14.83	Poor match (MW 119)	33
443	6.50E-05	0.011094	19.03	2-hydroxy-2-methylpropionate hydroxymethyl	59
451	8.09E-08	0.00011	19.18	Poor match (MW 114)	43
460	1.26E-05	0.003449	19.42	Dimethyl disulphide	94
933	1.52E-05	0.003454	34.82	2 methyl acetate1 butanol	43

Figures 23 and 24 illustrate examples of two VOCs which showed good matching with the NIST library: neopentane (peak 208) and dimethyl disulfide (peak 460). Consistent with these findings, compounds from the same groups including short chain alkanes and sulfides were noted to be increased in bacteria cultured on agar plates (as shown in chapter 3).

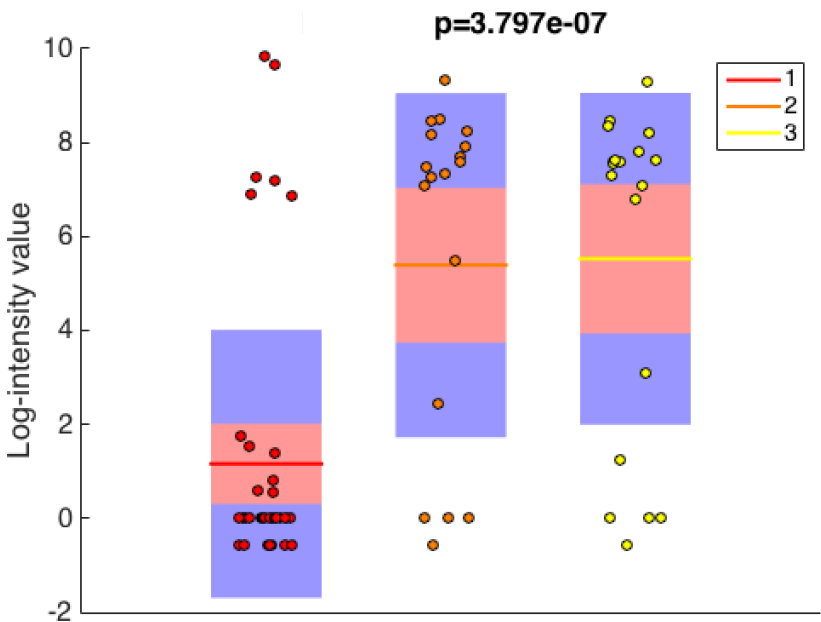


Figure 23: Induction of neopentane in headspace of BEAS2B cells stimulated with *S. pneumoniae*, *H. influenzae* or medium control. Following groups: 1 = Medium treatment. 2 = *S. pneumoniae* infection 3 = *H. influenzae*. Data analysed by Kruskal Wallis test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

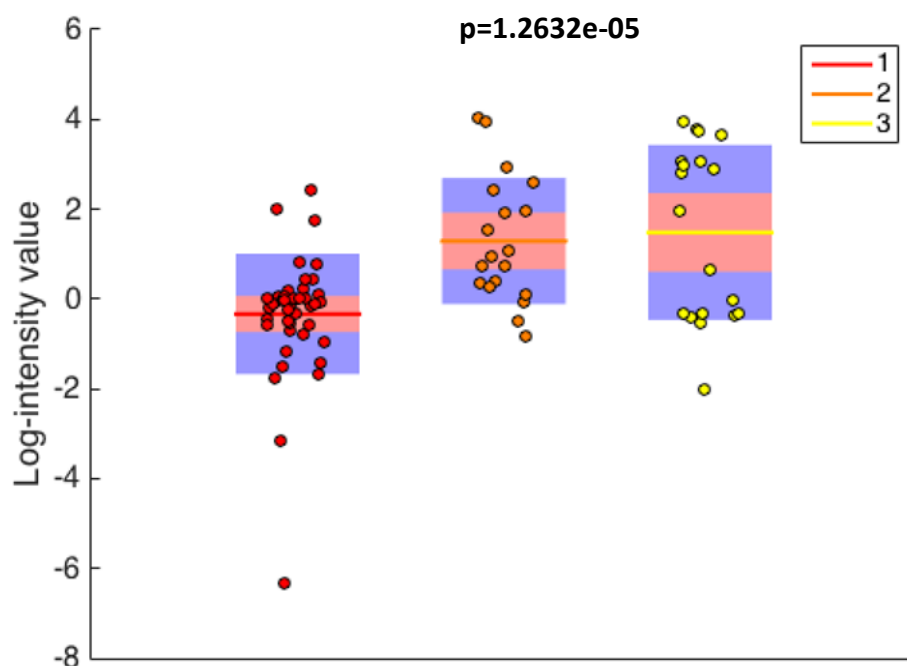


Figure 24: Induction of dimethyl disulfide in headspace of BEAS2B cells stimulated with *S.pneumoniae*. *H.influenzae* or medium control. Following groups. 1 = Medium treatment 2 = *S.pneumoniae* infection 3 = *H. influenzae* Data analysed by Kruskal Wallis test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

4.4.7 Combined bacterial infection analysis

I next performed a combined analysis of experiments using *S. pneumoniae* and *H. influenzae* infection to improve power and evaluate whether specific 'bacterial' VOC signatures could be identified in the headspace of BEAS2B cells. There were 4 FDR corrected VOC metabolites that were identified with good separation and classification (table 8). This was associated with an AUC of 0.72, indicating moderate predictive value (Figure 25).

Table 8: Discriminatory VOCs between combined bacteria *S. pneumoniae* and *H. influenzae* stimulated cells versus medium treated control cells

<i>peak</i>	<i>p values</i>	<i>q values</i>	<i>fold change</i>	<i>Retention time</i>	<i>compound</i>	<i>M/z base peak</i>
52	0.000117	0.03999	-1.7533	4.99	Methyl nitrite	30
156	1.51E-06	0.002064	-1.3083	8.43	3-methyl-1-butyne 1,4 Pentadiene	67
460	1.72E-05	0.011718	-2.3067	19.42	Dimethyl disulphide	94
702	7.55E-05	0.03434	-1.1712	28.45	4 methyl2heptanone	43

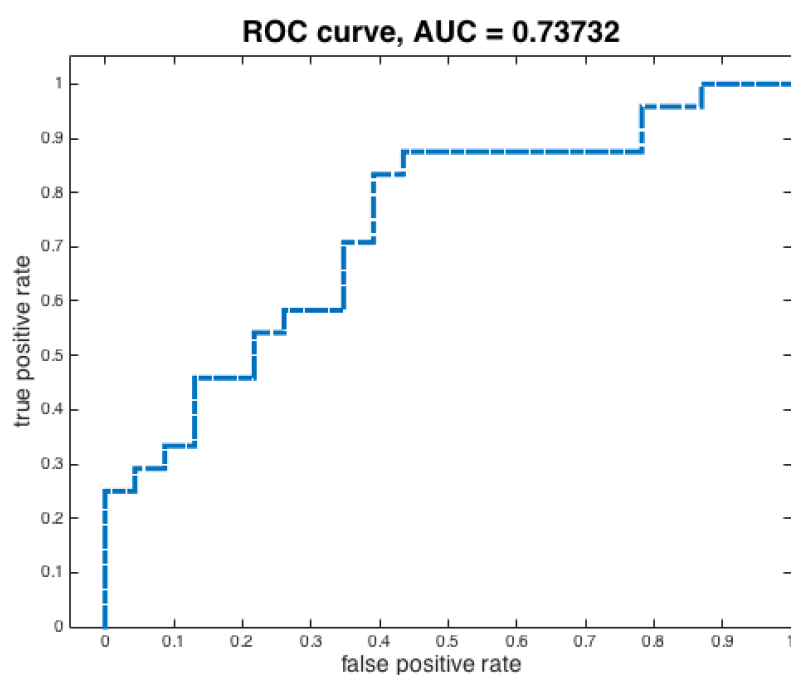


Figure 25: ROC curve showing predictive value of VOC analysis for differentiation between bacteria and medium stimulated cells.

Figures 26 and 27 graphically illustrates the increase in two of these compounds dimethyl disulphide and methyl nitrite in bacterially stimulated cells versus medium treated cells.

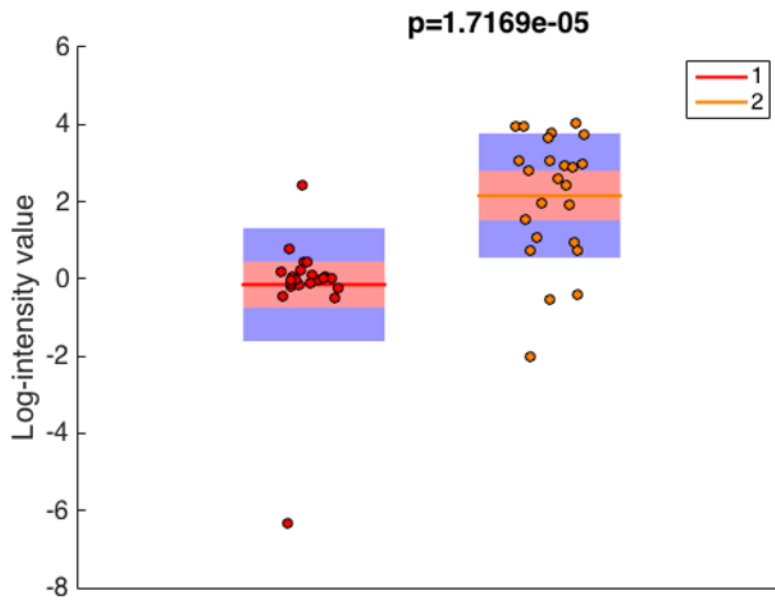


Figure 26: Induction of dimethyl disulfide in cells infected with bacteria. Following groups: 1=media, 2=Bacteria with media and cells. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

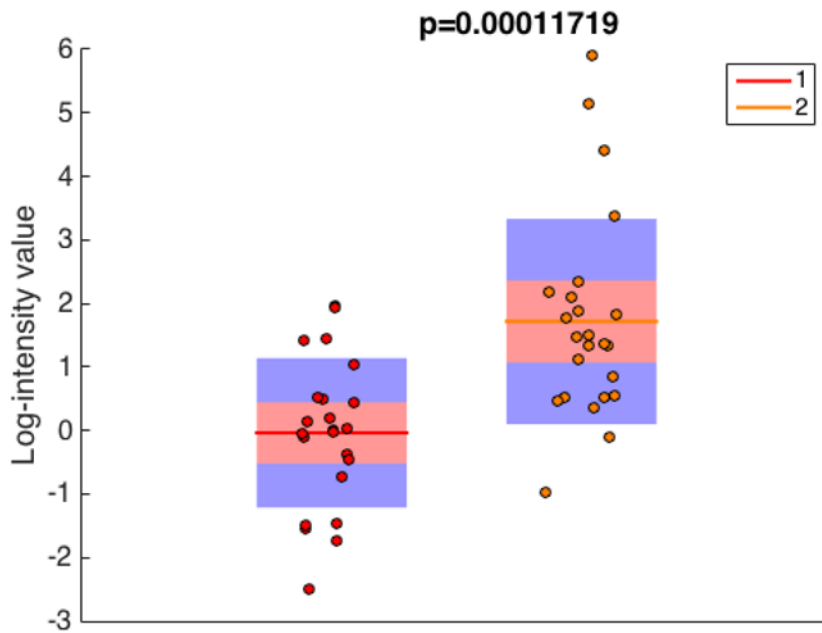


Figure 27: Induction of methyl nitrite in cells infected with bacteria. Following groups: 1=media, 2=Bacteria with media and cells. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

As previously shown, there is a small component or effect of the presence of cells on VOC profiles, which may be impacting the separation of profiles (AUC of 0.74).

After correction for this confounder, logistic regression analysis showed an improved AUC of 0.89 indicating a good predictive test (Figure 28). Seven VOCs were identified that were statistically different between bacterially stimulated and medium treated cells (table 9).

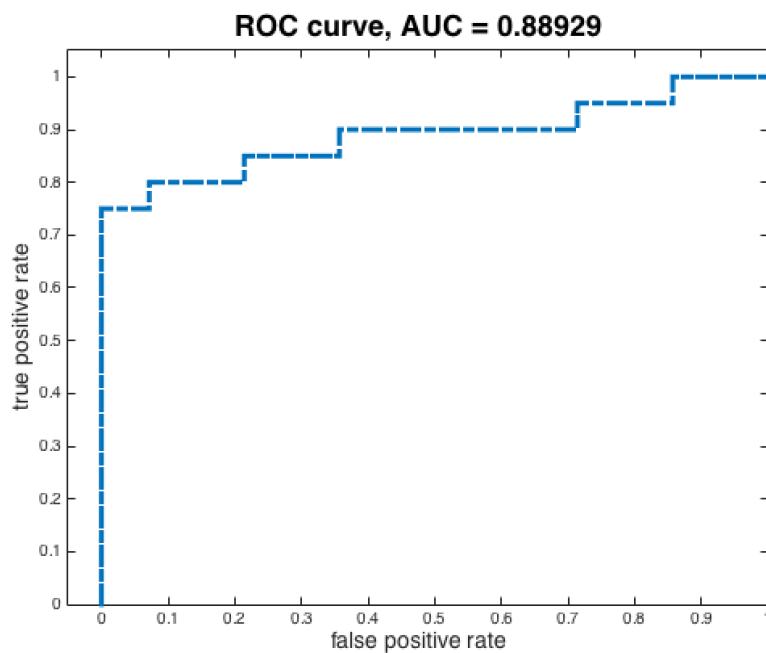


Figure 28: ROC curve showing separation between control (media) and media containing bacteria

Table 9: Discriminatory VOCs between bacterially stimulated and medium treated cells after correction for the presence of cells

<i>Peak</i>	<i>p value</i>	<i>Q value</i>	<i>fold change</i>	<i>Retention time</i>	<i>Compound</i>	<i>M/z base peak</i>
190	1.24E-05	0.003398	3.68	9.82	Poor match (MW 106)	57
197	3.97E-05	0.007749	1.8297	10.07	2methyl-2propanol	59
205	1.71E-05	0.003883	3.4853	10.33	Cyclobutylamine	43
208	2.62E-07	0.000357	5.7201	10.39	Neopentane	57
263	1.37E-06	0.000933	2.5322	12.30	3-ethoxy-3-methyl-2butanone	59
335	8.34E-06	0.002847	1.8996	14.83	Poor match (MW 119)	33
451	5.47E-06	0.00249	5.1305	19.18	4,5dimethyl-1,3-dioxol-2-one	43

Figures 29 and 30 illustrate examples of compounds found to be increased in bacterially stimulated versus medium treated cells. Again, similar classes of compounds were identified as observed in experiments with bacterial culture on agar plates (chapter 3). These included short alkanes, alcohols, sulphides and aldehydes.

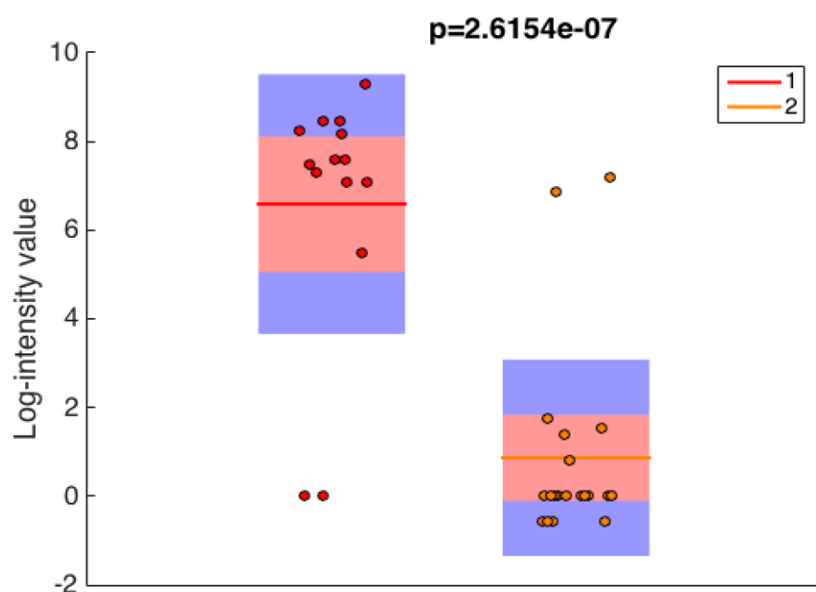


Figure 29: Induction of 4, 5 dimethyl 1,3 dioxyl-2 in headspace of BEAS2B cells stimulated with *S. pneumoniae*. *H. influenzae* or medium control. Following groups. 1= bacterial infection 2 = Medium treatment. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

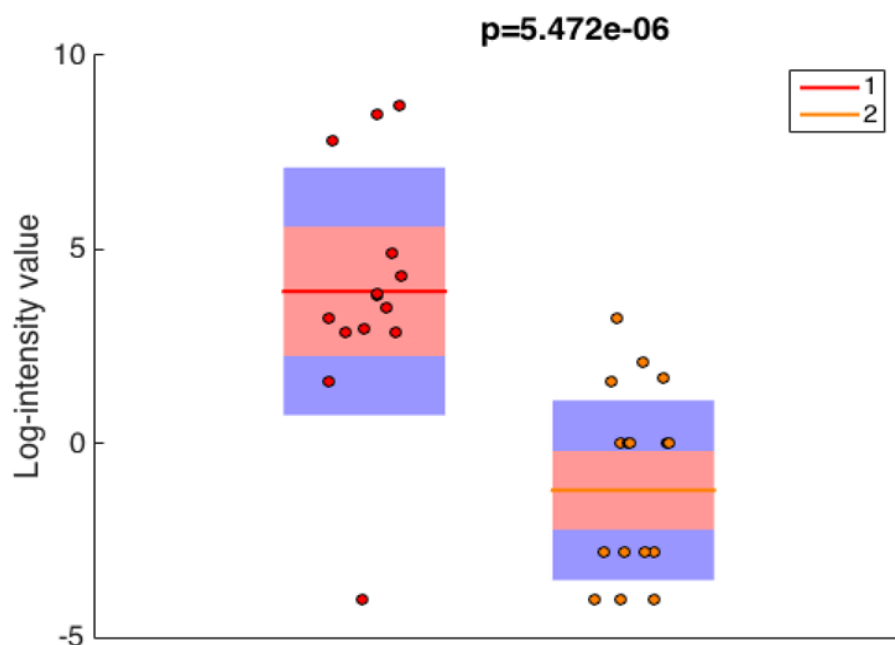


Figure 30: Induction of neopentane in headspace of BEAS2B cells stimulated with bacteria or medium control. Following groups. 1 = bacterial infection. 2 = Medium treatment. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

4.5 Discussion

The *in vitro* work conducted here has indicated groups of VOC compounds that are altered following infection of airway epithelial cells by common COPD-relevant bacteria and viruses. As discussed in chapter 3 the limitations of Tedlar bags have been highlighted and subsequent VOCs identified should be recognised with this in consideration. VOCs that were common to both viral and bacterial infection included alcohols such as 2-methyl 2-propanol, sulphides such as dimethyl disulphide and fatty acids such as acetic acid. VOCs that were specific to viral infection included long chain alkanes. RV infection of BEAS2B cells led to production of similar VOC compounds that have been reported previously including acetone, dimethyl sulphide and phenol¹⁰⁷. The data additionally indicates that RV infection led to increased production of decane (a long chain alkane). Production of decane derivatives have been shown previously with viral infections such as vesicular stomatitis virus¹¹⁰. Conversely, RSV infection of cells in the current study did not show an upregulation of decane. Dodeconate, an anion of dodecane, a 12 carbon membered alkane was observed upon RSV infection, while decane a 10 carbon membered alkane was seen with RV. This may suggest a similarity between both viruses in terms of VOCs produced. The production of these compounds may be linked to oxidative stress during infection which would be expected to occur with both RV and RSV infection¹⁴⁸. Spencer *et al* previous reported virus-related increases in fatty acid synthesis¹⁴⁹. Subsequent fatty acid degradation of the increased levels of lipids could be a potential source of the increased longer chain alkanes detected with RV infection. The reason for the difference between RV and RSV in upregulation of decane remains unclear and requires further study.

RSV infection of BEAS2B cells showed an increase in the following VOC groups: fatty acids, alcohols, aldehydes and ketones. These groups are distinct from those observed with RV. Although long chain alkanes were not observed with RSV, short chain alkanes, long chain fatty acids and aldehydes (greater than 10 carbon chain compounds, c10) were noted. The differences between VOC production related to RV and RSV infection may be due to various differences in the composition, replication and interaction of the virus with the host cell. Further studies are needed to determine this.

Bacterial infection of BEAS2B cells led to production of aldehydes, short alkanes (c4-7), alcohols, ketones and sulphide compounds. These compounds were also identified when the same bacteria were cultured on agar plates (butanol, 2-methyl 2-propanol,

neopentane and dimethyl disulphide). The detection of these compounds from both bacteria cultured on agar plates and infected epithelial cells suggests that these VOCs may be produced directly by bacteria or by the process of bacterial replication rather than by the host response.

There were a number of compounds that were produced by both bacteria and viruses in my experiments (see summary in Figure 31). These included alcohols, sulphides and phenols. Interestingly, 1-alcohols are produced through β - or α -oxidation of fatty acid derivatives through acetyl-CoA¹³⁸. Fatty acids are precursors for a variety of important building blocks such as phospholipids, sphingolipids and sterols. The degradation of fatty acids yields a high amount of ATP and their detection in my studies could represent an energy substrate facilitating pathogen growth. Saturated long chain fatty acids c26–c37 have also been previously shown to be required to increase viral activity and infectivity of human cytomegalovirus¹⁵⁰. Sulphur compounds such as dimethyl sulphide is likely to be a cleavage product of proteins such as cysteine and methionine amino acid residues which contributes to volatile emissions during proteolysis¹⁵¹.

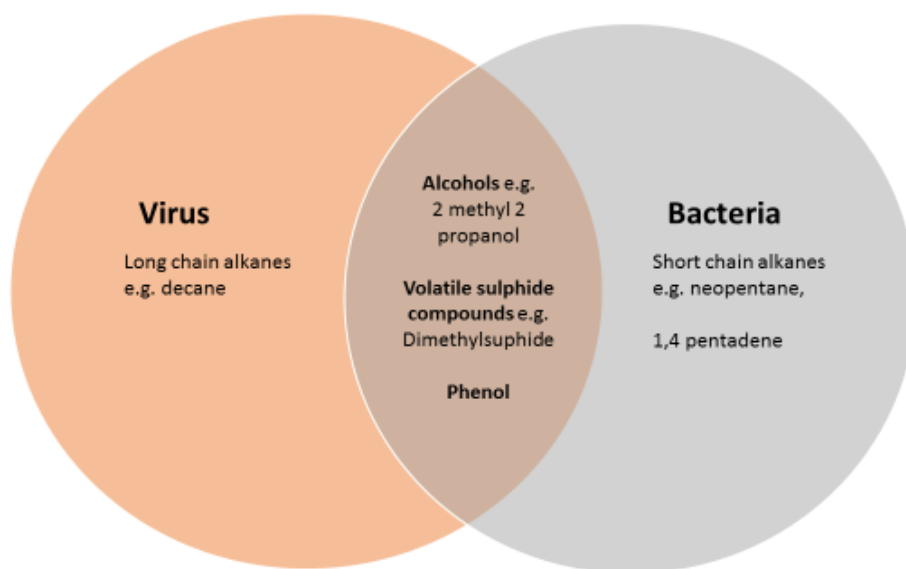


Figure 31: Venn diagram summarising VOC compounds produced by virus and/or bacterial infection in BEAS2B experiments.

As shown in Figure 31, other compounds were specific to either virus or bacterial infection. A number of pathogen or host response related differences could account for these observations. For example, viruses require a host for replication and do not have their own metabolism when compared to bacteria¹⁵². In contrast to bacteria, viruses dramatically modify cellular metabolism in their host upon entry into cell¹⁴⁶. Bacteria are able to generate their own energy substrates¹³⁹ and this may explain why short chain alkanes were upregulated by *S. pneumoniae* and *H. influenzae*.

Viral infection process is quite complex and involves multiple chemical reactions in the host cell. During the infection process, the viral envelope fuses with a vacuole membrane, and viral RNA replication takes place in the nucleus¹¹¹. The viral RNA and nucleoprotein (NP) are released into the cytoplasm for the RNA-dependent RNA transcription of the complementary positive-sense cRNA. The fusion is achieved by influx of protons into the viral envelope and decreasing pH due to acidic conditions in the endosome¹⁵³. Viral infections and subsequent changes in protein synthesis with minor alterations to a cell's genome can lead to differences in its VOC production¹⁵⁴. Cell infection, initiates signal transduction and protein expression pathways simultaneously, so an even larger downstream VOC production effect may be seen^{110,111}.

As discussed previously in chapter 3, limitations of not assessing VOC production between commensal and pathogenic bacteria identified. Non-pathogenic viruses VOC production in a controlled in vitro environment, too could have been assessed to show the potential feasibility and efficacy in a future clinical in vivo study.

4.6 Conclusion

Experiments in airway epithelial cell cultures have identified virus- and bacteria-specific VOC signatures. Specifically, alkane compounds could be useful as predictors of viral (long-chain variants) or bacterial (short-chain variants) infection. Importantly, many of the VOCs identified in the bacterial airway epithelial cell experiments shown here were also observed when the same bacteria were cultured on agar plates, indicating that they arise from the presence or replication of bacteria rather than as a result of the host immune response. These studies provided a framework for subsequent *in vivo* investigations of VOCs in human breath detailed in chapters 5 to 7. The feasibility of

using VOCs as markers to establish cause of exacerbations has been demonstrated. Using the established GC-MS methodology and analytical methods described here, there is promise that measurement of VOCs could potentially help to establish aetiology of infection in human subjects.

.

5 VOC collection from human clinical studies of healthy subjects

5.1 Introduction

This chapter will discuss human breath collection and the analysis undertaken in healthy human cohorts. A ‘healthy’ non-smoking control human subject cohort group, with no medical comorbidities, was initially established to run a clinical trial to assess the clinical efficacy of a cold therapeutic. The author provided the clinical lead and helped conduct this study and thus was able to collect breath from subjects experimentally infected with rhinovirus (RV16).

The common cold is one of the most frequently encountered infectious pathogens in man¹⁵⁵. Despite the benign nature of the illness in the majority of healthy people, it can cause pathology in people with respiratory diseases and remains a significant economic burden on society¹⁵⁵ in persons with chronic respiratory disease, such as asthma, COPD and cystic fibrosis due to the spreading of the virus to the lower respiratory tract¹⁵⁶. Rhinovirus leads to an increase in consultations with clinicians, increased absence from school and work, and subsequently causes loss of earnings. In such patients there exists an increased likelihood of complications such as secondary bacterial infections². The clinical benefits of identifying markers to help distinguish bacterial and viral exacerbations is an exciting prospect and one that will be explored in this chapter. The symptoms patients experience can be both similar and vague, thus the assessment by clinicians can be subjective and subsequent treatment can be variable.

The author was a co-investigator in a clinical trial evaluating the efficacy of an over the counter spray (ColdZyme), for the prevention of the common cold. The ColdZyme solution contained glycerol, water, cod trypsin, ethanol (<1%), calcium chloride, tris and menthol¹⁵⁷. This study was funded by Enzymatica, a Swedish based company that was keen to show evidence of efficacy of their spray as a cold prevention. A previous study in Sweden with much smaller numbers designed by the company showed a reduction in cold symptoms and duration of infection with reduced viral loads¹⁵⁷. The function of the spray is to create a physical barrier at the back of the throat thereby decreasing viral replication and stopping the development of symptoms. It is thought that epithelial cell layers in oral and nasal cavities form a physical and innate immune barrier against

bacteria and viruses¹⁵⁸. However, viruses can infect the mucosal cells in this area resulting in cold symptoms. Hence, strengthening the natural epithelial barrier may be of benefit in order to inhibit viral entry into host cells by free virus particles in the throat mucosa. ColdZyme® Mouth Spray (ColdZyme) is a medical device designed to challenge against the common cold. It is designed to deposit a viscous solution containing primarily glycerol and cod trypsin onto the throat, which, thereby, reduces the probability of catching a cold and helps shorten the duration of a cold by forming a thin protective barrier on the oropharyngeal mucous membrane.

ColdZyme has been shown to have an effective broad-spectrum impact against common cold viruses in previous studies,¹⁵⁹ thereby reducing viral efficacy. The current study was a trial experimental infection using RV16 infection, evaluating the efficacy of ColdZyme in combating the cold symptoms. This provided an opportunity to collect exhaled breath from healthy subjects before, during and after acute infection with RV16. The aim was that the experimental infection with RV16 in healthy subjects, with subsequent breath sampling, will identify VOCs related to acute viral exacerbation, and may identify VOCs similar to those identified *in vitro*. The subjects all being healthy non-smoking volunteers would minimise confounding factors such as medications, medical comorbidities and poor health, thus increasing the likelihood of detected volatiles being directly attributable to infection.

5.2 Method

5.2.1 Study Design

The study design employed was a double-blind, randomized, placebo-controlled, parallel-group study. Ethics was obtained from a local Research Ethics Committee (REC No. 15/ES/0112). Subjects were randomized 1:1 to use either ColdZyme, or a placebo mouth spray. The objectives of this study were to investigate the effect of ColdZyme on rhinovirus replication in the oropharyngeal mucosa and on common cold symptoms *in vivo*. The participants in the study were deemed to be healthy and this was assessed and judged through the investigator's thorough and detailed medical history and physical examination in the screening visit. Further inclusion criteria for the study were the willingness and ability to complete the study and the perception of having had at least one cold per year. Use of ColdZyme or placebo was also documented.

For the main part of the study, subjects attended the hospital/clinic site six times as identified in the above diagram. Prior to the first visit, at screening, subjects completed medical questionnaires and had a blood serology test. This was to

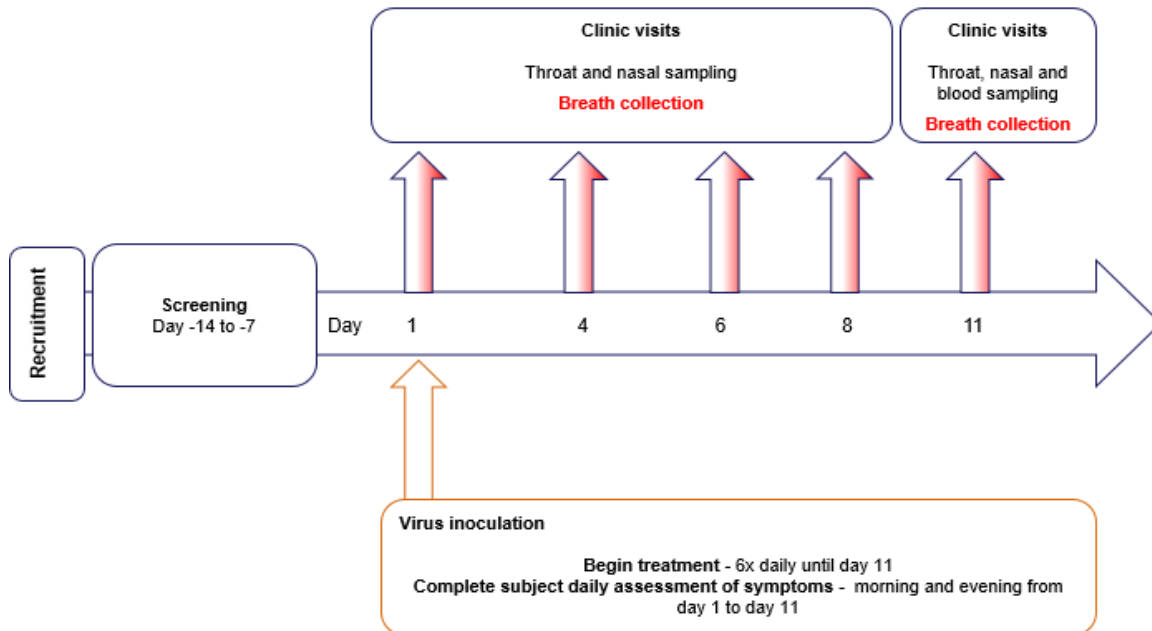


Figure 32. Overall schematic for RV challenge trial in human volunteers. At screening, healthy volunteers were recruited into the study based on having no underlying respiratory disease and sero-negativity for RV16. At Day 1, they were infected with 100 TCID₅₀ RV16 via nasal inhalation, and sampled on days 4, 6, 8 and 11 for virus load in nasal lavage and throat swab, Exhaled breath VOCs and symptom score using a modified Jackson scale.

ensure patients were well, virus naïve and healthy prior to exposure to RV16. Previous exposure to RV16 was assessed through blood serology (RV16 antibody presence) to ensure subjects that took place had not been immunologically exposed in the past and therefore should develop infection if exposed/inoculated. The degree of response to inoculation from the host is less predictable but through measuring the presence of viral load in other fluids such as throat swab and nasal lavage, the presence of active virus can be at least quantified and thus be used, at least partially, as a surrogate for potential to cause acute infection.

Volunteers were excluded if they met any of the following criteria identified on the day of screening: smoker during the last 12 months; any cold symptom within the last month such as sore throat, sneezing, rhinorrhea, malaise, nasal obstruction or cough; presence of serum neutralizing antibodies against rhinovirus 16 at screening (blood

samples were collected and were screened at National Heart and Lung Institute, Imperial College London, UK); active allergic rhinitis, asthma or chronic obstructive pulmonary disease during the previous year; positive responses to cat or dog allergens (if subject was likely to come in contact with the specific pet) and/or dust-mite allergens in a skin prick test at screening; nasal disease (e.g., nasal polyposis, significant septum deviation, chronic rhinosinusitis, etc.); females who were pregnant, breastfeeding or intended to become pregnant during the study; active autoimmune disease during the previous year; evidence or history of drug or alcohol abuse; use of any prescribed or non-prescribed medication (except for contraceptives, paracetamol and ibuprofen) within 2 weeks prior to the first administration of investigational product until the end of study; use of any over-the-counter cold prophylaxis products such as vitamin C, zinc or Echinacea within 1 month prior to the first administration of investigational product until the end of study; and participation in any other clinical studies within 60 days. Subjects were free to withdraw their consent at any time without affecting future treatments.

5.2.2 Sampling and assessments

Techniques and samples collected from subjects are described below:

- Oropharyngeal swab – a soft nylon flocked rolled along the tonsil and throat area to dislodge and collect cells. This sample is used to detect presence of virus in the throat.
- Nasal lavage - gently wash both nostrils with 2.5mL (1 teaspoon) of saltwater (normal saline). Viral PCR was undertaken to look for virus and if required inflammatory cytokines in the sample, to check if subjects infected with the cold virus through the presence of viral load and how the body responds to it.
- Breath collection- Subjects were asked to perform a deep nasal inhalation followed by complete exhalation via their mouth into a secure GastroCHECK steel breath bag (250 mL) (Bedfont Scientific LTD, and Ardmore Healthcare LTD, UK) through a mouth piece with one way valve. Subjects asked to breathe through steel bag at least 2 times to use own breath to clean bag through. As patient breathes through bag for third time, breath was captured by capping both ends. Total duration of collection is less than 20 seconds.

Previous work within the group has demonstrated that there is a loss of trace VOCs from breath bags that impair interpretation but steel bags appeared to have the best performance in reducing loss of trace aldehydes¹³¹

Breath samples at five time points (baseline and day 4, 6, 8 and 11) were collected from each of the 22 subjects and analysed by GC-MS. Additionally, viral loads from nasal lavage and oropharyngeal swabs were measured. Of the 22 subjects, an additional breath sample was collected at 28 days post-infection for eleven of the subjects as our initial preliminary analysis revealed changing of peak VOC levels during acute infection towards pre-existing levels. Breath, throat swabs and nasal lavage samples were collected at visits on day 4, 6, 8, 11 and 25 post-infection (Figure 33).

5.2.3 Experimental RV16 infection

These subjects were all experimentally infected with RV16 via an intranasal inoculation technique as previously described^{36,160}. Briefly an atomiser (De Villis where) was used to deliver a total of 250 microliters of RV16 (100 TCID₅₀) into both nostrils. Any remaining solution was applied to the nasal cavity by the volunteer's index finger.

5.2.4 Measurements of virus load

Viral Polymerase Chain Reaction (PCR) analysis was performed on nasal lavage and throat swab samples as previously described^{36,160}. This analysis was undertaken at the St Marys Hospital National Heart Lung Institute at Imperial College London, Respiratory Medicine, by Dr Tata Keadze and Dr Mike Edwards.

5.2.5 VOC transfer to thermal desorption tube

Breath was then entrained onto a thermal desorption tube using an airflow pump at 50mL/min for 3 minutes with the aim of analysing the majority of samples within 24 hours to maximise VOC detection and prevent degradation.

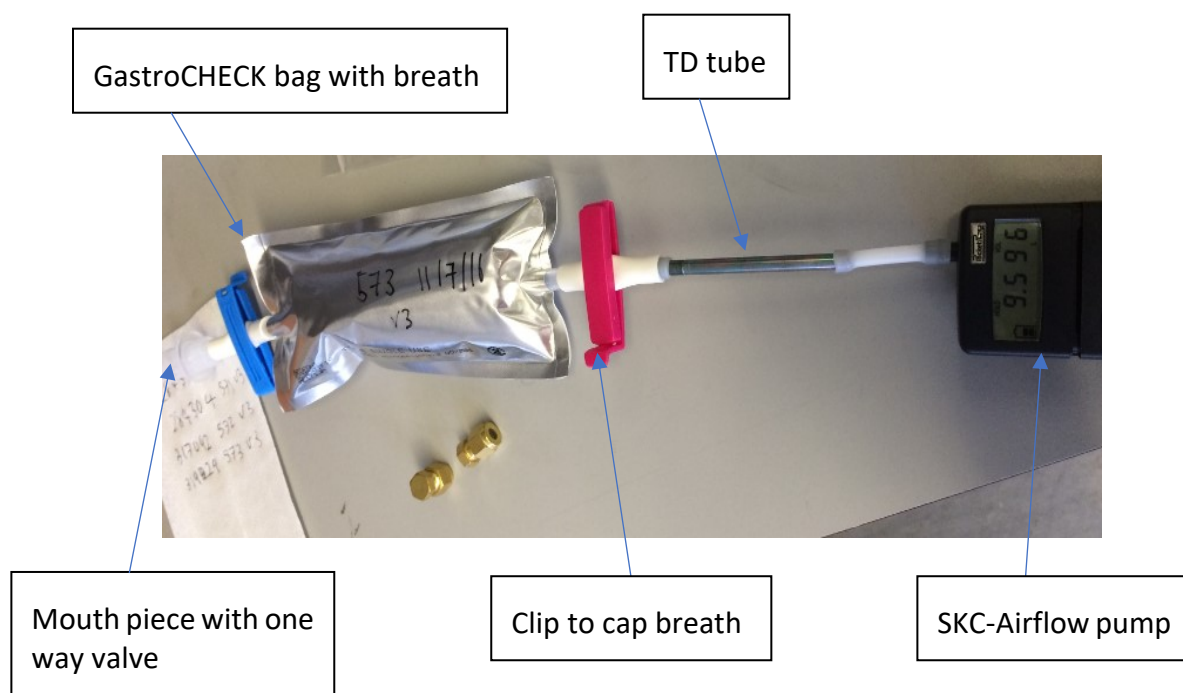


Figure 33. Human breath collection. Showing breath transfer from breath collection bags onto a TD tube using an airflow pump.

Prior to sampling, tubes were conditioned at 325°C for 40 minutes in a stream of nitrogen passed through a hydrocarbon trap (Supelco, US) using a Markes International TC-20 tube conditioner, operated according to manufacturer's recommendations. This was to ensure tubes were clean and free of contaminant VOCs prior to usage. With each TD GC-MS run on, a 'cleaned' TD tube was run every ten runs on the GC-MS to allow for rapid assessment and quality check assurance.

5.2.6 GC-MS analysis

Exhaled breath samples concentrated onto TD tubes were analysed using GC-MS. The TD tubes were desorbed using a Markes TD-100 TD unit (Markes International Ltd, Llantrisant, UK) using a two stage desorption programme, applying a constant flow of helium at 50 mL min⁻¹. In the primary desorption stage, TD tubes were dry-purged for 1 min and heated at 280 °C for 10 min. In the secondary desorption stage, the cold trap (U-T12ME-2S, Markes International Ltd, Llantrisant, UK) was rapidly (99 °C min⁻¹) heated to from 10 °C to 290 °C. VOCs were transferred from the TD unit to the GC by means of a capillary line heated at 140 °C. GC-MS analysis was performed using an Agilent 7890B GC with 5977A MSD (Agilent Technologies Ltd, Santa Clara, USA) equipped with a ZB-642 capillary column (60 m × 0.25 mm ID × 1.40 µm df; Phenomenex Inc., Torrance, USA) with helium used as the carrier gas (1.0 mL

min⁻¹ flow rate). The GC column temperature settings and methodology remained stable throughout my PhD as stated in the previous method sections.

5.2.7 Data analysis

A toolbox was developed through MATLAB and R packages, which was used to align VOC peaks and normalise the data. Subsequently levels of each VOC were plotted over time and trends observed. Those VOCs that have the highest peak to baseline variance were ascertained and plotted. Clustering analysis was undertaken (comparative VOC) and mean trajectories calculated and displayed.

Pattern recognition analyses were performed on the processed spectral data from processed GC-MS breath samples using MATLAB. Initially, principal component analysis (PCA) of the GC-MS spectral data was performed to visualize group clustering, that is, overall similarities and differences between spectroscopic profiles, and to identify any abnormalities or outliers within the data set. Bi-cross validation was used to ensure that the principal components captured systematic variation was not attributable to noise¹⁶¹. Subsequently, a supervised, multivariate analysis method, maximum margin criterion (MMC) was employed to simultaneously maximize the variation between groups, whilst minimizing intra-group differences¹⁶².

Cross-validation with quadratic classification was used to assess the predictive capacity of the models. The spectral profiles of each breath sample and subject were withheld from the dataset one at a time. The discriminating components, that is, linear combinations of metabolic features that separate classes in a mathematically optimal way, were derived based on the remaining data. The withheld profiles were then projected onto the discriminating space and assigned to the class to which they had the smallest distance via the quadratic classifier. The cross-validation was repeated until the spectral profiles of all breath samples were predicted. The classification accuracies (that is, confusion matrices) and predicted variance were used to assess the performance of the supervised multivariate models.

Potentially discriminatory metabolites were selected using “training” profiles by one-way analysis of variance with a liberal *p*-value threshold of 0.05 (ANOVA¹⁶³, *p*-value < 0.05) and their collective capacity to discriminate between classes was tested on withheld (“test”) profiles using the above multivariate modelling strategy. Additionally, one-way ANOVA with a FDR of either 0.1 or 0.05, to account for multiple hypothesis testing¹⁶⁴,

was applied to identify metabolites that individually (irrespective of other metabolites) discriminate between classes, based on suitably adjusted threshold p -values. Typically, the adjusted p values corresponding to an FDR of <10% would be significantly < 0.05.

5.3 Results

5.3.1 Assessments of virus load

The nasal and oral viral PCR results showed that after inoculation there was a peak viral load at day 4 with subsequent return to baseline by day 11. Each subject's viral load was used and the mean was calculated and plotted for viral loads for both nasal and oropharynx. Greater viral levels were noted through the nasal viral load measurements at all time points comparing that to levels of virus from the oropharynx.

The overall study for the efficacy of ColdZyme contained 88 participants, (44 placebo and 44 ColdZyme subjects) but by the time ethics were amended to allow for breath collection in the study, 23 subjects were recruited in total that allowed for breath collection as well as the other tests carried out. 11 female and 12 male subjects with an age distribution 20- 66 years were infected with RV16 and had their breath collected, as well as the nasal and oropharyngeal swabs. The data below include volunteers received both placebo or ColdZyme, and who had confirmed successful experimental infection by either positive serology or increases in nasal or oral virus load.

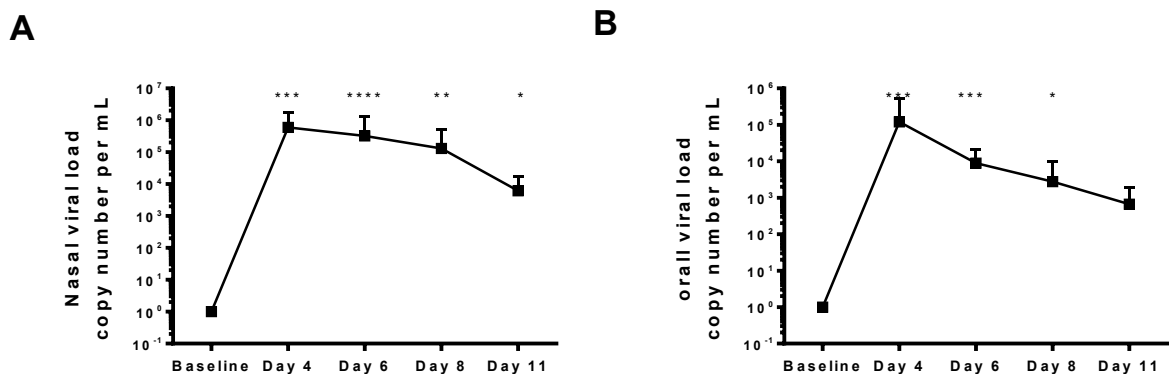


Figure 34: Mean viral loads in (A) nasal passage and (B) oropharynx following experimental RV infection in 21 healthy subjects. Two subjects that did not have detectable virus load in either nasal lavage or oral swab were determined to be uninfected. Nasal lavage performed at baseline and 4 time points post RV infection and measured using quantitative PCR. Where a virus load of 0 copies was determined, this was converted to a value of 1. MEAN and SEM shown * p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001 versus baseline.

5.3.2 Identification of VOCs in infected subjects

Using R analysis software (later reconfirmed with MATLAB) and NIST library, initially fifty VOC profiles were shown to have an increase in levels from baseline to days 4-8 (during acute infection). This was then followed by what appears to be a return to near baseline levels of these compounds. Twelve of the fifty peaks were found to have good matching (forward and reverse matching circa 800 and favourable RI indices) with NIST library when assessing compound identification. Figure 35 shows an example of two of the twelve VOCs (4,7-dimethyl-undecane and decane) which showed this pattern. The line graphs show the individual breath samples and subsequent levels of each VOC from baseline to day 25.

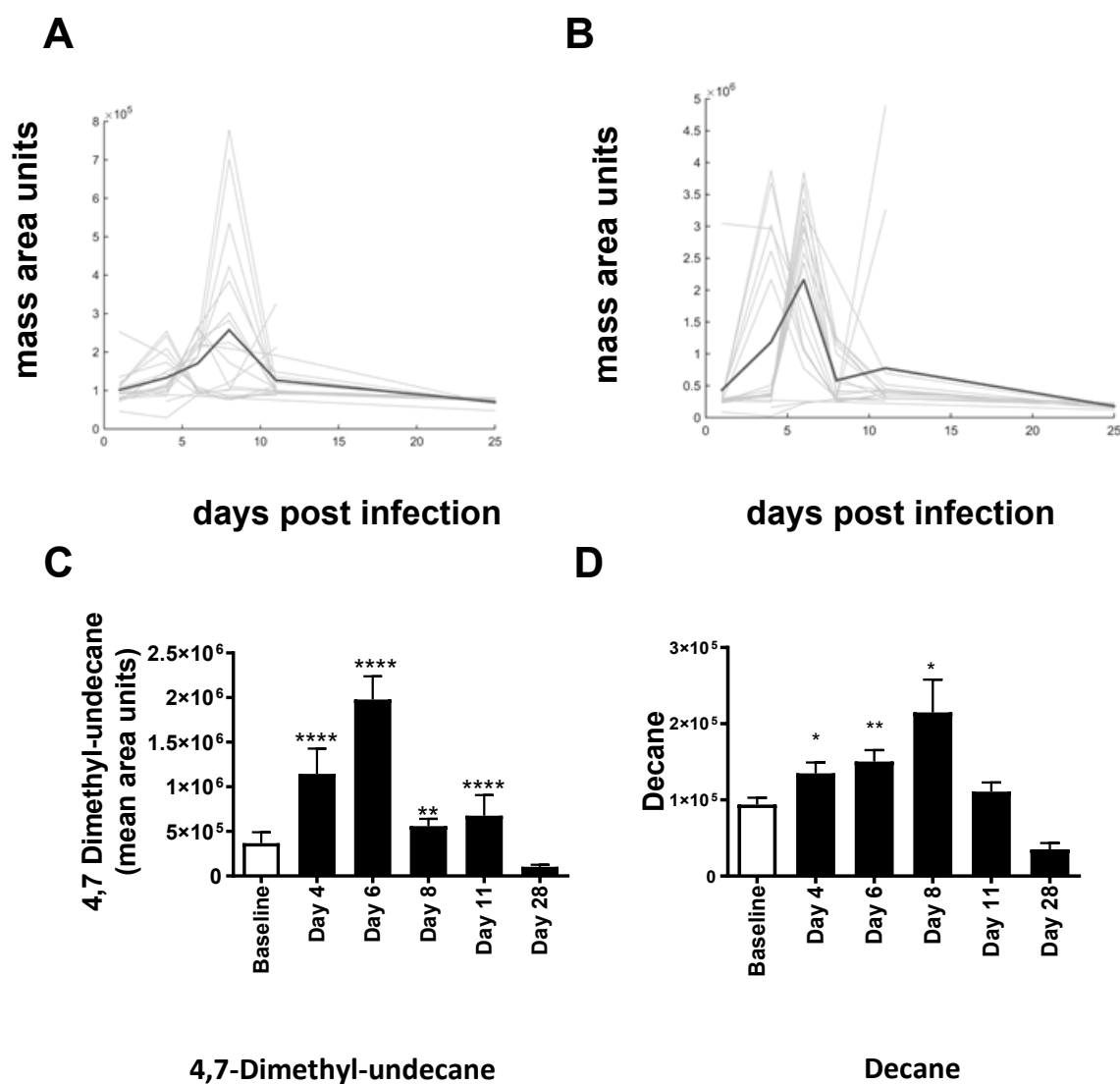


Figure 35: Line graph and box plots showing VOC induction in RV16 infection of healthy human 2 of 12 VOCs shown; 4,7 dimethyl undecane (A,C) and decane (B,D). The other 10 show very similar pattern of VOC level expression related to days post infection. One way ANOVA analysis shown * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ versus baseline.

Other VOCs following similar pattern and good matching through NIST library include 3 ethyl hexane, 2,4 dimethyl heptane, hexadecane, 3 methyl pentane, 3,8 dimethyl undecane, 4 methyl dodecane, cyclohexane and 4,6,6 pentamethyl heptane.

Clustering analysis undertaken using MATLAB which identified that the 288 VOC metabolites displayed 14 potential patterns that VOC metabolites followed (Figure 36).

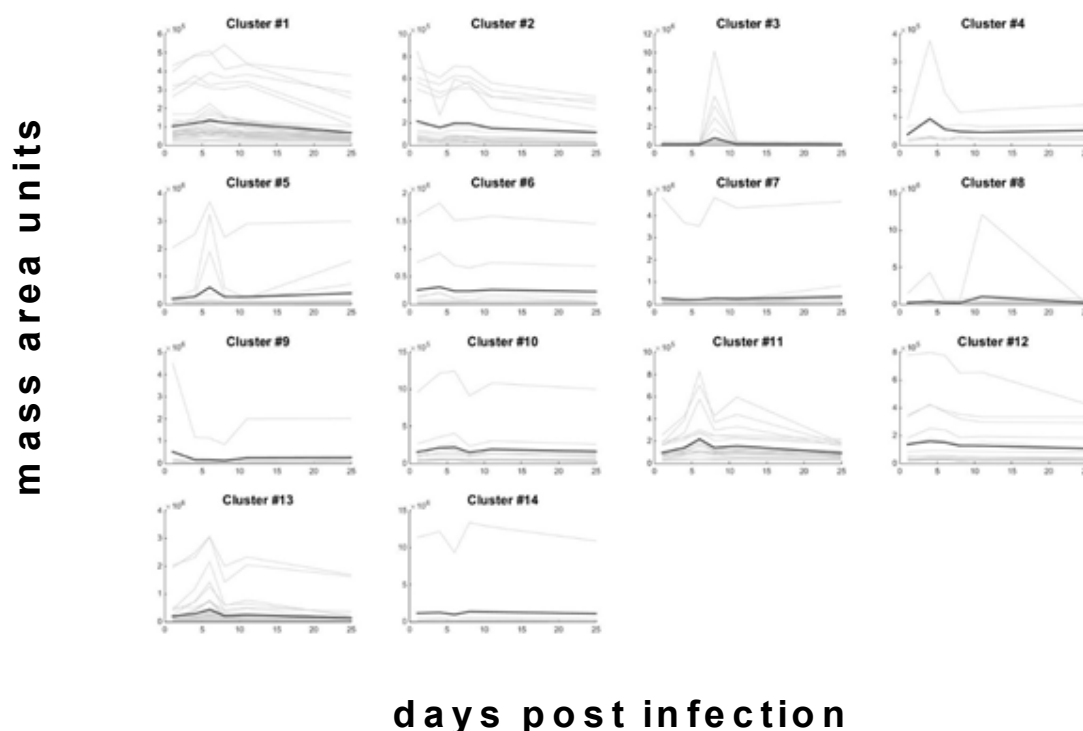


Figure 36: shows the 14 cluster patterns displayed by the differing VOC peaks in human subjects infected with RV. Overall 288 VOC Peaks identified.

Using ANOVA and FDR correction with subsequent cross referencing with NIST, VOC metabolites showing significant changes from baseline were identified (table 10). Approximately 55 peaks were identified through analysis that showed significant compounds, identified tentatively when possible through NIST (forward and reverse matching >750 with highest probability offered by NIST and mass spectral analysis).

Table 10: Top 15 most significant VOCs increased in RV infection in healthy humans shown

Peak	Retention time	FDR adjusted p value	Compound	m/z base peak
15	9.4	1.0E09	2-Propanol	45
223	25	7.46E-09	1,4-dimethyl-Benzene	91
200	26.2	7.99E-09	1,3-dimethyl-benzene	91
60	40	9.56E-09	2,3,7-trimethyl-4-Octene	69
87	39	1.59E-08	3,8-dimethyl-undecane	71
236	24.6	2.14E-08	Ethylbenzene	91
234	14.4	3.05E-08	Cyclohexane	84
6	33.1	3.79E-08	4,7-dimethyl-undecane	71
20	31.5	4.30E-08	5,7-dimethyl-undecane	71
26	13.1	5.18E-08	2-Butanone	43
42	32.9	1.24E-07	Dodecane	71
25	32.7	1.52E-07	Undecane	57
237	39.5	2.22E-07	3,8-dimethyl-undecane	71
173	38.7	3.64E-07	Hexadecane	71
167	29.0	0.000497164	Decane (105 additional source)	57
66	42.6	3.84E-07	Unknown (MW 162)	162
259	35.6	6.90E-07	Unknown (MW 166)	151
29	30.4	7.01E-07	1,2,3-trimethyl-benzene	105
19	31.3	8.79E-07	4,7-dimethyl-undecane	71
160	37.5	9.08E-07	Unknown (MW 270)	107
249	35.8	9.66E-07	Unknown (MW 166)	70
106	36.1	1.32E-06	Unknown (MW 170)	57
108	40.6	1.32E-06	Unknown (MW 268)	71

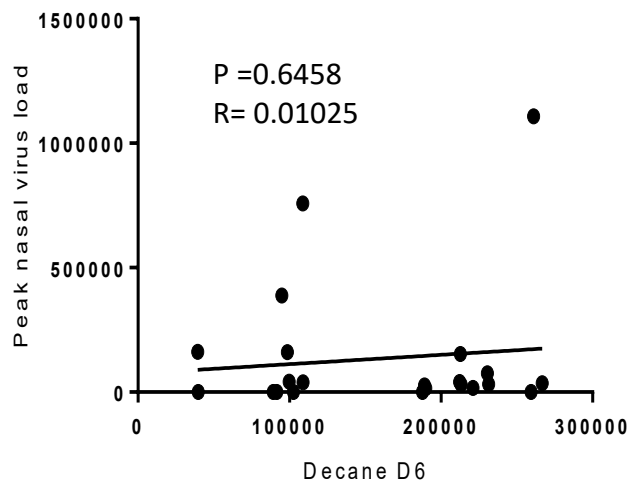
An assessment of the effect of cofounders using MATLAB package and Kruskal-Wallis analysis with false discovery rate (FDR) correction was undertaken. The ColdZyme study was initially set up as a double blinded study, ColdZyme spray vs placebo (control) investigations. Analysis undertaken using two way analysis of variance. No effect of ColdZyme spray and placebo on metabolite expression was noted, with no FDR corrected values less than 0.05. Therefore, all data shown represents a combination of patients receiving both active drug and placebo.

5.3.3 Association of VOCs and virus load

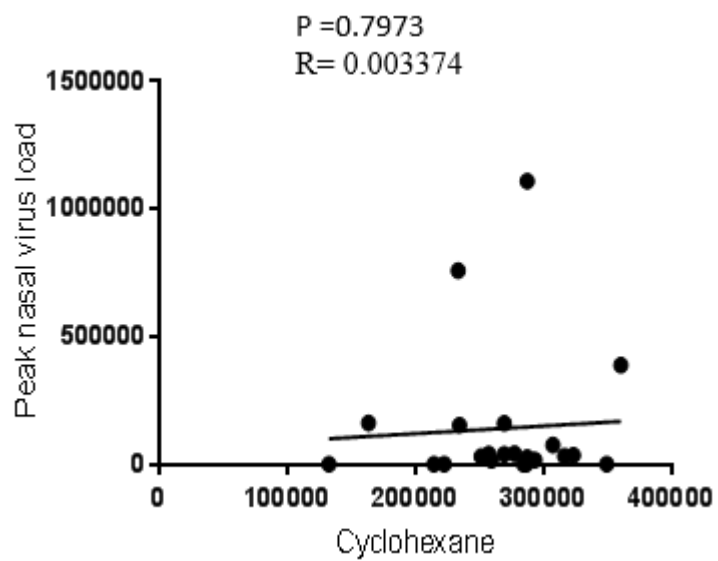
Regression analysis was undertaken for the fifteen most significant VOC metabolites (three shown in Figure 38), to assess the potential relationship between viral load (nasal

PCR results were chosen as they were more consistent than oral virus load, and looking at previous literature, the nasopharyngeal sample site is more consistent and representable than oral sampling site). There was no correlation found between viral loads and any volatiles; Figure 37 is an example of three compounds analysed.

A



B



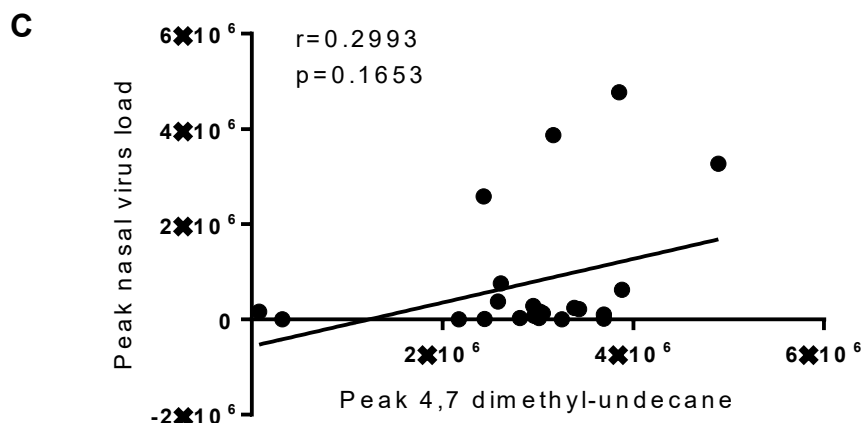


Figure 37: Three examples of regression analysis undertaken to establish the potential relationship between nasal viral load and elevated VOC. A Decane, B cyclohexane and C 4,7 dimethyl-undecane. No correlation found.

The best correlation was seen with benzyl methyl sulphide which shows a trending towards correlation but still does not achieve statistical significance.

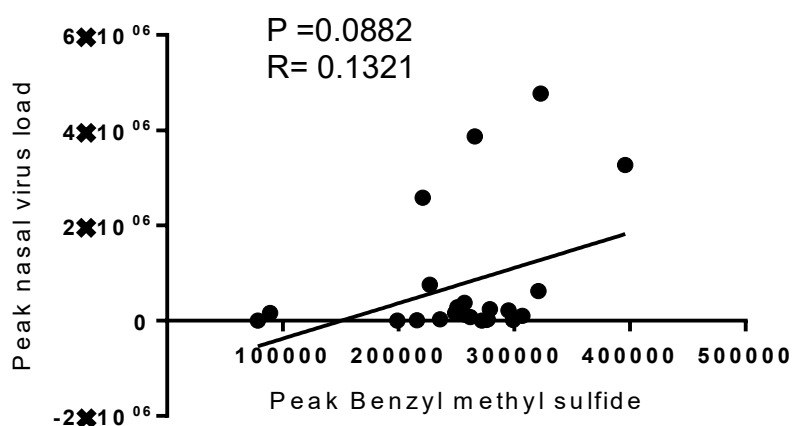


Figure 38: Benzyl methyl sulphide regression analysis undertaken to establish the potential relationship between nasal viral load and elevated VOC

5.3.4 Associations of VOCs and Symptom score

Symptom scores between male and female subjects assessed between male and female subjects $n=23$. There was $N=12$ male and $N=11$ female subjects. Symptoms were ranked between 0-5 for each day for the following; sneezing, rhinorrhea, nasal obstruction, sore/scratchy throat, cough, malaise, hoarseness and chilliness for each

day and then the total for each day subsequently summated and the total numbers then analysed. There was no statistically significant difference with regards total symptoms between female and male subject's p values 0.75 for mean, 0.78 median and p value 0.61 (established using t-Test).

Mean male symptom scores trended slightly higher than female scores but no statistical significance achieved (see Figure 39). Confidence interval of the mean demonstrated in pink, blue bars represent confidence interval for whole data. Overlap of the confidence interval of the mean visually correlates with a lack of significance between the groups.

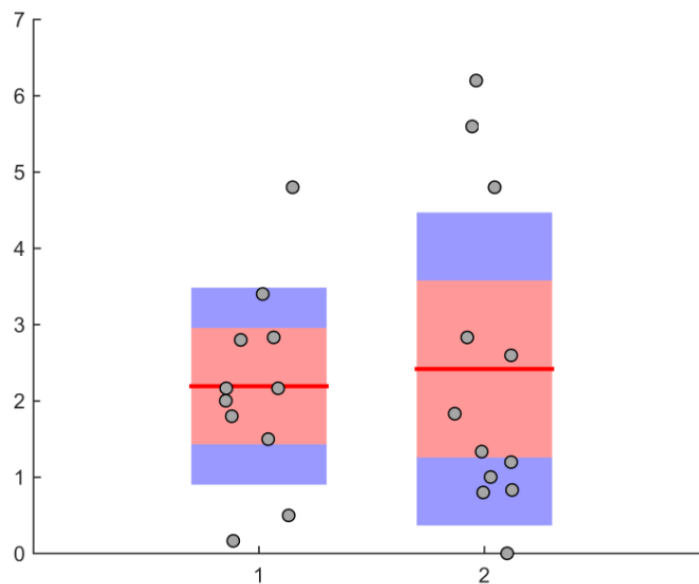


Figure 39: Total symptom scores following RV infection. Groups: 1= female 2= male. Data analysis using t-Test. Red line is the mean.

Regression analysis undertaken assessing correlation between VOCs detected and symptom score and no correlation was detected between as shown in Figures 39 and 40 below between peak VOCs and peak and total symptom score.

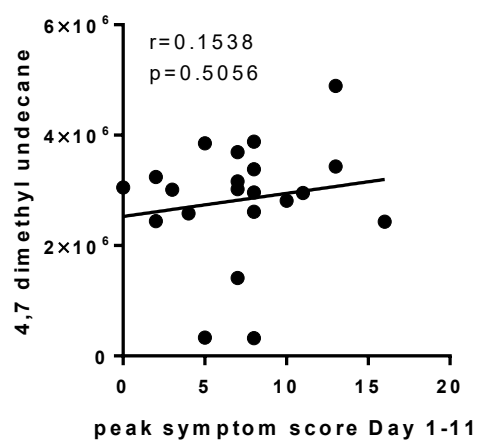
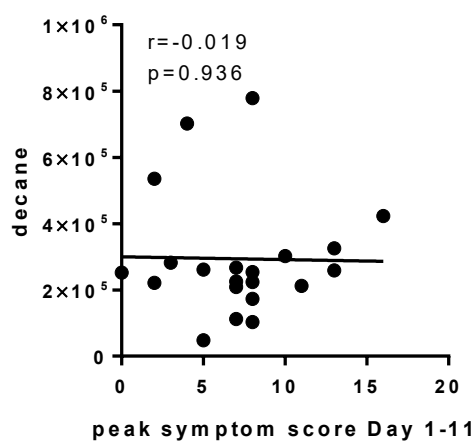
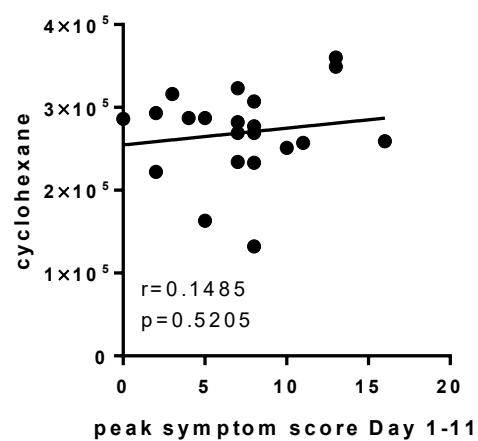
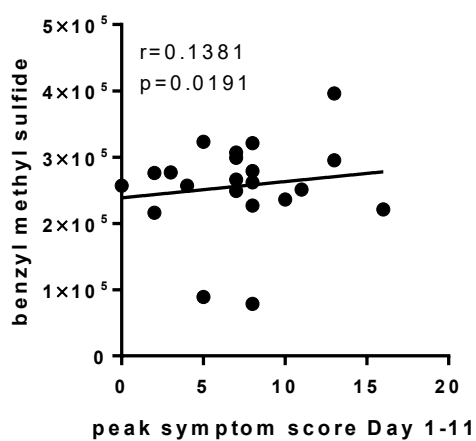
A**B****C****D**

Figure 40: Relationships between peak VOC and peak symptom score (A) 4,7 dimethyl undecane (B) decane (C) cyclohexane (D) benzyl methyl sulphide.

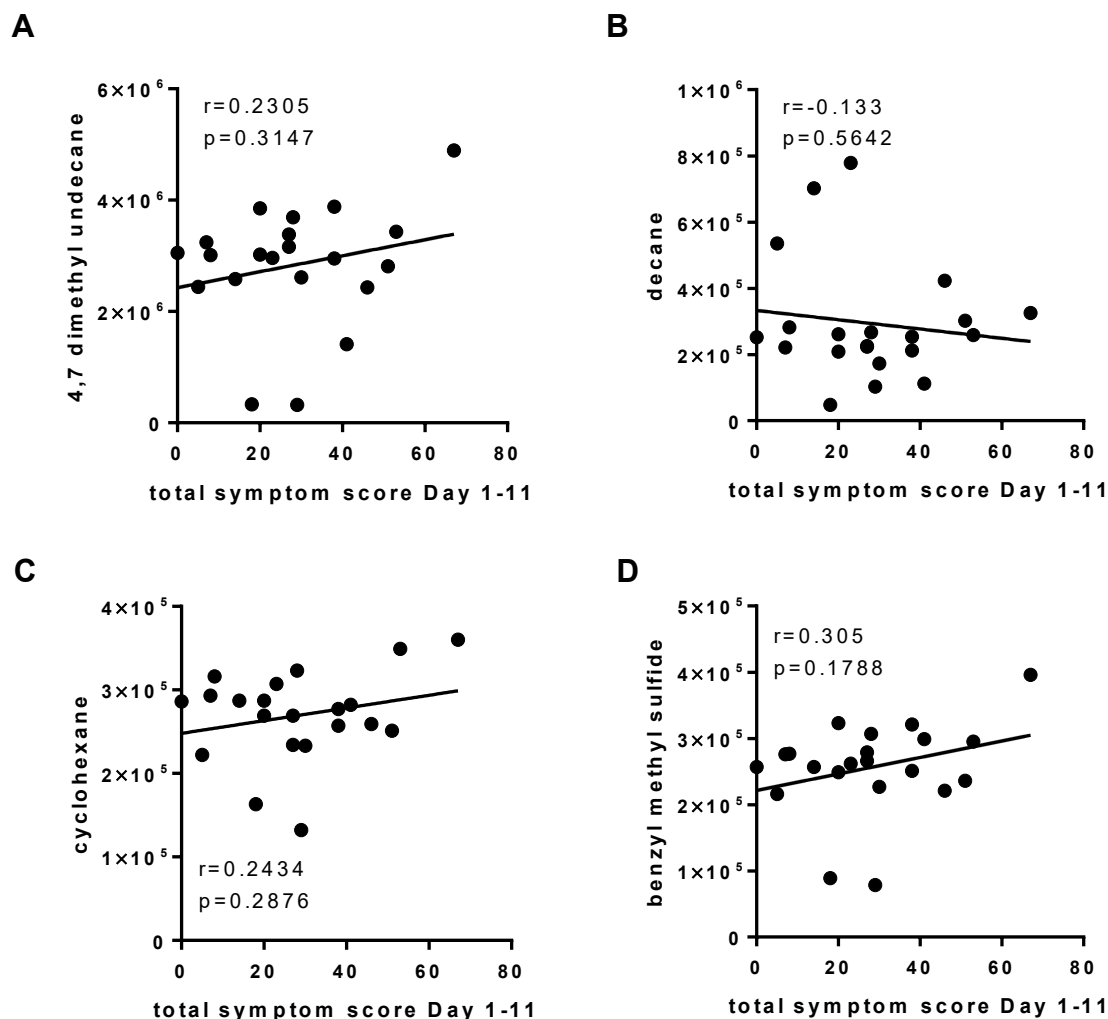


Figure 41: Relationships between peak VOC and total symptom score (A) 4,7 dimethyl undecane (B) decane (C) cyclohexane (D) benzyl methyl sulphide.

5.4 Discussion

The overall study result assessing the efficacy of coldzyme is classified and the thesis is not at liberty to present the end result of the study (data subject to confidentiality claims from the study sponsor). The study was unblinded and through regression analysis we found the coldzyme spray had no effects on VOCs of interest and thus enabled the use of combining both active and placebo cohorts in analysis. As the ethics to allow for breath collection allowed for sampling only after halfway through the study process the small numbers $N=22$ mean any results and conclusions must bear this in mind. This study is the first of its kind and certainly does show some promise towards identifying VOCs towards viral RV infection. Human breath in healthy subjects has been

found to comprise of different 872 VOCs¹⁶⁵. Phillips *et al.* have noted long chain/branched aliphatic compounds; predominantly alkanes/alkenes and alcohols (specifically alkanes and methylated alkanes) were previously reported as biomarkers of live attenuated influenza viral vaccination in breath¹¹⁰. The breath analysis of RV infected human subjects undertaken in the current study has also shown an increase in alcohols including 2-propanol, and long chain alkanes e.g. decane, dodecane and undecane post inoculation with RV with a subsequent fall in levels after presumed recovery from infection.

Viruses are non-living entities and do not have their own metabolism when compared to bacteria¹⁶⁶. Viruses dramatically modify cellular metabolism in their host upon entry into a cell. Conversely, bacteria are able to generate their own energy substrates and hence short chain alkanes seen¹³⁹. However, during viral infection host cells metabolise high levels of glycoproteins and other lipid material to provide the necessary energy required for virion production¹⁴⁶. Therefore, one could expect that the VOCs observed during a viral infection are qualitatively different from that of bacteria. Previously, Spencer *et al* showed a virus related increase in fatty acid synthesis¹⁴⁹. Subsequent fatty acid degradation of the increased levels of lipids could be a potential source of the longer chain alkanes noted in RV infection. In both the *in vitro* and human RV infection model analyses, long chain alkanes were noted to be increased as well as sulphides containing group compounds (e.g. dimethyl disulphide and benzyl methyl sulphide). Sulphur containing compounds such as dimethyl sulphide have been detected in both bacterial and viral infections¹¹⁰. Dimethyl sulphide is likely a product of protein degradation involving amino acids such as cysteine and methionine, which contribute to volatile emissions during proteolysis¹⁵¹. As any viral infection is a dynamic process, it is reasonable to assume that complex biochemical events determining the fate of different proteins as infection progresses. Though the exact mechanisms by which VOCs are produced and released are not always clear, it is likely that they are products of biochemical reactions that occur during an infection of a host cell. As these biochemical reactions change during the course of an infection, the VOCs from these reactions would change as well.

The total symptom score for each subject was calculated and the total mean scores trended to be higher in males than females although not found to be statistically significantly different. Gender differences in the perception of signs and symptoms colds

have previously been interrogated, the phenomena of 'man flu'. Even after adjustment for other variables, men were significantly more likely to 'over-rate' their symptoms in comparison with the clinical observer than were women. Rather than artefactually exaggerating gender differences in morbidity, differing thresholds for perceiving and reporting symptoms are thought to produce underestimates of the gender differences in morbidity¹⁶⁷. The results of my RV human study seem to corroborate this.

5.5 Conclusion

A pattern of several VOCs with potentially decane and decane derivatives were found to increase from baseline levels in volunteers experimentally challenged with RV-16 produced as a response to infection. Previously, in Chapter 4 decane was also found to increase in levels in RV infection of BEAS2B cells in culture, and hereabouts has been seen to increase from baseline during RV16 infection in previously healthy subjects although be it in a small number. Together this consistently suggests that decane or decane containing VOCs are associated with the host response to RV-16. Decane is an alkane produced by lipid peroxidisation and its biological significance will be discussed later in chapter 6. Male subjects' symptoms scores trended to be higher than the female counterparts within the study but there were no significant differences noted between the two genders. Although there was no correlation between peak viral loads, peak and total symptoms scores and peak VOC metabolite levels we did show presence of virus through nasal and oropharyngeal loads, symptom scores trending to increase post inoculation. These however are limited evidence of active viral infection from when breath collected and subsequent extrapolation of results from this chapter should take this into account.

6 London COPD cohort

6.1 Introduction

The aim of this study was to investigate the diagnostic accuracy of the VOCs noted *in vitro* and in RV infected healthy human subject's breath samples to differentiate between bacterial and viral exacerbations in a cohort of patients with COPD monitored prospectively in the community.

The results identified through the RV human challenge study provided relative promise within a homogeneous population and the aim of the next component of this thesis was to see if these observations were transferable to a more heterogeneous 'real-world' cohort of subjects experiencing naturally occurring infections in the community.

The cohort is a very well characterised group of COPD subjects who have been followed for many years. Data from symptom score, FEV1, blood markers has been collected annually, during stable state and during a COPD exacerbation. This cohort was set up originally by Professor Wedzicha whom moved from the Royal Free Hospital, London to the Royal Brompton Hospital, London in 2011, and she continues to oversee various prospective longitudinal studies investigating COPD progression and COPD exacerbation markers⁵⁰.

With the promising results noted in the RV16 infection of healthy subjects, the present study then focused on this more heterogeneous population of patients with COPD. The aim was to investigate if breath analysis was able to differentiate between causes of exacerbation of COPD and also whether there were general VOC markers when unwell, comparing to when subjects were well or stable. For the purposes of the present study, subjects were followed with data collection over approximately 18 months (November 2016 – March 2018) and breath analysed using GC-MS and PTR-ToF-MS instruments to investigate specific VOCs or groups of VOCs could be used to establish the infective cause of the observed exacerbations.

6.2 Method

The London COPD Cohort is an established group of around 200 well characterised COPD patients (Inclusion criteria: COPD spirometry displaying $FEV_1/FVC < 0.7$ (clear obstructive spirometry) and ages ≥ 45 . Exclusion criteria: <10 pack year smoking, significant other respiratory disease - REC Reference 09/H0720/8). The aim was to recruit 88 patients and follow them over a 14 to 18 month period. This sample size is based on previous similar studies evaluating biomarkers and aetiology of COPD exacerbation¹⁶⁸. Through our *in vivo* rhinovirus inoculation trial conducted in healthy subjects, a sample size of $n=22$ would have 80% power to detect a 30% increase in VOC concentration assuming a standard deviation of 198730.4, using a paired t-test with a 0.05 two-sided significance level.

6.2.1 Study design

All subjects followed had samples at baseline, when in a healthy (control) stable state, aiming to follow through exacerbation (within 24h of symptoms day 0, 7) and after subsequent recovery samples if possible (day 14, 35). Patients were reviewed for a duration of 18 months (with three to six month regular baseline visits undertaken). The subjects were trained to report for assessment and sampling before initiating treatment (antibiotic and corticosteroids), although this was followed with variable success. Once a patient reported feeling unwell, scoring 2 or more symptoms on diary card for more than 2 consecutive days, a clinical visit was organised to assess their status as soon as possible (otherwise they continued to be seen at 3-6 monthly intervals as baseline stable scheduled visits).

Longitudinal cohort study

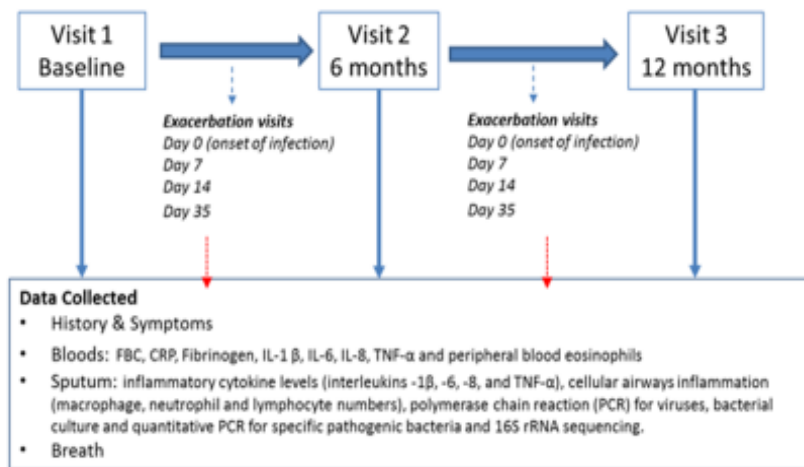


Figure 42: Study design for investigating VOCs as novel biomarkers in exacerbations of COPD.

6.2.2 VOC breath collection method

Breath was collected using steel breath bags as described in the previous HRV-A16 human infection challenge study, described earlier in this chapter using GastroCHECK bags (Bedfont Scientific LTD and Ardmare Healthcare LTD, UK). The method of collection was the same as described previously in the ColdZyme study methodology section 5.2.2. As per the previous methodology, breath was transferred to TD tubes and analysed on GC-MS and PTR-TOF-MS ideally within 24 hours as per Markar *et al*¹³¹ to reduce degradation of VOC markers and allow the best opportunity to detect VOCs of interest.

6.2.3 VOC mass spectrometry analysis

With the COPD cohort breath samples, breath was also entrained from steel bags and analysed through a second analytical platform PTR-TOF-MS, direct online analysis, as well as GC-MS using established methodology. GC-MS analysis methodology has been described previously in chapter 3 and 4. Optimal conditions for VOC identification and quantification were defined through work undertaken within the VOC group at St Mary's Hospital, London, as per the previously described experimental workflow, dedicated to method optimization under breath-relevant conditions¹⁶⁹. Experiments were conducted under breath-relevant conditions (100% humid air), and with optimised settings;

temperature 110°C, pressure 2.30 mbar and voltage 350 V, resulting in an E/N (field density ratio, E being the electric field strength and N the gas number density) of 84 Td (1 Townsend = 10⁻¹⁷ V cm²). Reactivity was inspected using H₃O⁺, NO⁺, and O₂⁺ as primary ions. The results from previous work within the group has shown that a relatively low (<90 Td) E/N reduces fragmentation enhancing sensitivity and identification capabilities, particularly in the case of aldehydes using NO⁺, where a 4-fold increase in sensitivity is obtained by means of drift voltage reduction¹⁶⁹.

6.2.4 Samples other than breath collected

Other than breath, data collected included severity of airflow obstruction (Forced Expiratory Volume [FEV1]), smoking status, blood markers of infection CRP/WCC, sputum- for bacterial culture (bacteria positive or not) and sputum for viral PCR analysis and CAT scores (symptoms). Blood serology was stored as well as sputum for further possible testing of inflammatory cytokines such as IL-1, IL-6, IL-8 and TNF alpha for other potential markers (systemic) inflammation for collaborative findings. Initially, sputum was processed and stored as per protocol in appendices 8.2. At the end of the time period collection (18 months) all sputum was processed using viral PCR analysis for RV, RSV and influenza at the Division of infection and airway disease, St Mary's Hospital, Paddington by Dr Tata Keadze as per established well recognised PCR protocols. The PCR methodology including protein gel electrophoresis for RV, RSV and Influenza identified the presence or absence of virus rather than quantitative analysis.

6.2.5 Data analysis

Significant statistical support and advice was received from Dr Kirill Veselkov and Ivan Laponogov to analyse this complicated data set. Using MATLAB software, a bespoke model was used which undertook group separation based on the maximum margin criterion /support vector machine (MMC/SVM) multivariate model. Analysis was undertaken to establish discriminatory VOCs between: clinical exacerbation and baseline breath samples, bacterial positive (virus negative) sputum culture versus culture negative (bacteria and virus) exacerbation and baseline breath samples, viral positive and bacterial negative sputum culture and clinical exacerbation positive versus virus and bacterial negative sputum cultures in baseline subjects.

Pattern recognition analyses were performed on the processed spectral data from processed GC-MS breath samples using MATLAB. Initially, principal component

analysis (PCA) of the GC-MS spectral data was performed to visualize group clustering, that is, overall similarities and differences between spectroscopic profiles, and to identify any abnormalities or outliers within the data set. Bi-cross validation was used to ensure that the principal components captured systematic variation was not attributable to noise¹⁶¹. Subsequently, a supervised, multivariate analysis method, maximum margin criterion (MMC) was employed to simultaneously maximize the variation between groups, whilst minimizing intra-group differences¹⁶². Cross-validation with quadratic classification was used to assess the predictive capacity of the models. The spectral profiles of each breath sample and subject were withheld from the dataset one at a time. The discriminating components, that is, linear combinations of metabolic features that separate classes in a mathematically optimal way, were derived based on the remaining data. The withheld profiles were then projected onto the discriminating space and assigned to the class to which they had the smallest distance via the quadratic classifier. The cross-validation was repeated until the spectral profiles of all breath samples were predicted. The classification accuracies (that is, confusion matrices) and predicted variance were used to assess the performance of the supervised multivariate models. Potentially discriminatory metabolites were selected using “training” profiles by one-way analysis of variance with a liberal p -value threshold of 0.05 (ANOVA¹⁶³, p -value < 0.05) and their collective capacity to discriminate between classes was tested on withheld (“test”) profiles using the above multivariate modelling strategy. Additionally, one-way ANOVA with a FDR of either 0.1 or 0.05, to account for multiple hypothesis testing¹⁶⁴, was applied to identify metabolites that individually (irrespective of other metabolites) discriminate between classes, based on suitably adjusted threshold p -values. Typically, the adjusted p values corresponding to an FDR of <10% would be a significance of < 0.05.

The parametric (ANOVA/t-test) methods are relatively robust against deviations from normality, even in small sample sizes. In our case, the parametric tests are generally more powerful (due to the symmetrically distributed sample population) and are thus preferable. However, we did report more conservative Kruskal-Wallis (KW) analysis where possible to avoid this “symmetry” assumptions. More important in our case was to account for simultaneous comparative analysis of thousands of metabolic features. This was accounted by adjusting the p -values using Benjamini Hochberg procedure,

while controlling the rate of false discoveries among "statistically significant" metabolites at a certain percentage <5 or <10%.

In addition, the effects of smoking, degree of COPD as separated by the GOLD classification and the effect of medication with regards to distinguishing VOC markers and confounding effects on VOCs were interrogated. As previously stated, possible discriminatory VOCs were tentatively identified using the NIST library. Standards were not logistically possible on all compounds as such a vast array of compounds were identified. However through m/z and retention time analysis, this work can be undertaken with those compounds which were of potential importance or interest, in the future.

6.3 Results

The data collected from November 2016 to February 2018 and reported in this thesis was from 139 patients from within the COPD cohort at the Royal Brompton Hospital.

6.3.1 Overview of samples collected

The cohort data was collected from 91 male subjects and 48 female subjects. 96 of the subjects were ex-smokers, and 43 were current smokers with severity of obstructive airway diseases ranging from GOLD stage I-IV (full table of patient demographics found in appendices 8.1). The majority of subjects (82%) were from GOLD stage I & II. In total, 686 breath samples were collected. 686 breath samples run on the GC-MS. As with most clinical investigations, there was some loss of data with regards to inadequate patient compliance with clinical follow up, including the providing of sputum and undertake other investigations. This was minimised as much as possible through patient education at baseline visits and in any clinical interaction to reiterate the importance of getting in contact when unwell. There were 363 baseline breath samples and 323 exacerbations breath samples (total of 686 sample) run on the GC-MS instrument for analysis 278 completely matched breath/pcr/sputum samples were collected. A further 408 were unmatched due to two or more missing samples. The matched samples were from 85 subjects, of which there were 27 which had baseline and exacerbation samples to perform paired analysis. Of the 378 sputum cultures, 227 sputum plugs were located to perform viral PCR analysis.

6.3.2 Microbiological and virological assessments in sputum

A total of 378 sputum cultures were sent and analysed at the RBH microbiology. Subsequent analysis identified 229 upper respiratory tract flora (URTF); 146 Bacterial (20 *S. pneumoniae*, 28 *H. influenzae*, 52 *Pseudomonas spp*, 15 *Moraxella spp*, 14 Mixed (2 or more different bacteria), 12 *Staphylococcus spp*, 9 *Serratia spp* and 5 *Klebsiella spp* samples. 227 of the 378 sputum samples underwent viral PCR analysis; 32/227 found to be RSV positive and 22/227 RV positive. Of the 32 RSV positive sputum samples, over half had bacterial co infection 18/32 with positive microbiology noted. The majority of RSV positive samples (19/32) were from subjects at baseline. Of the RV positive subjects, 12/22 showed bacterial coinfection (pathogenic bacteria and virus positive) and the majority were taken from subjects having an exacerbation (14/22).

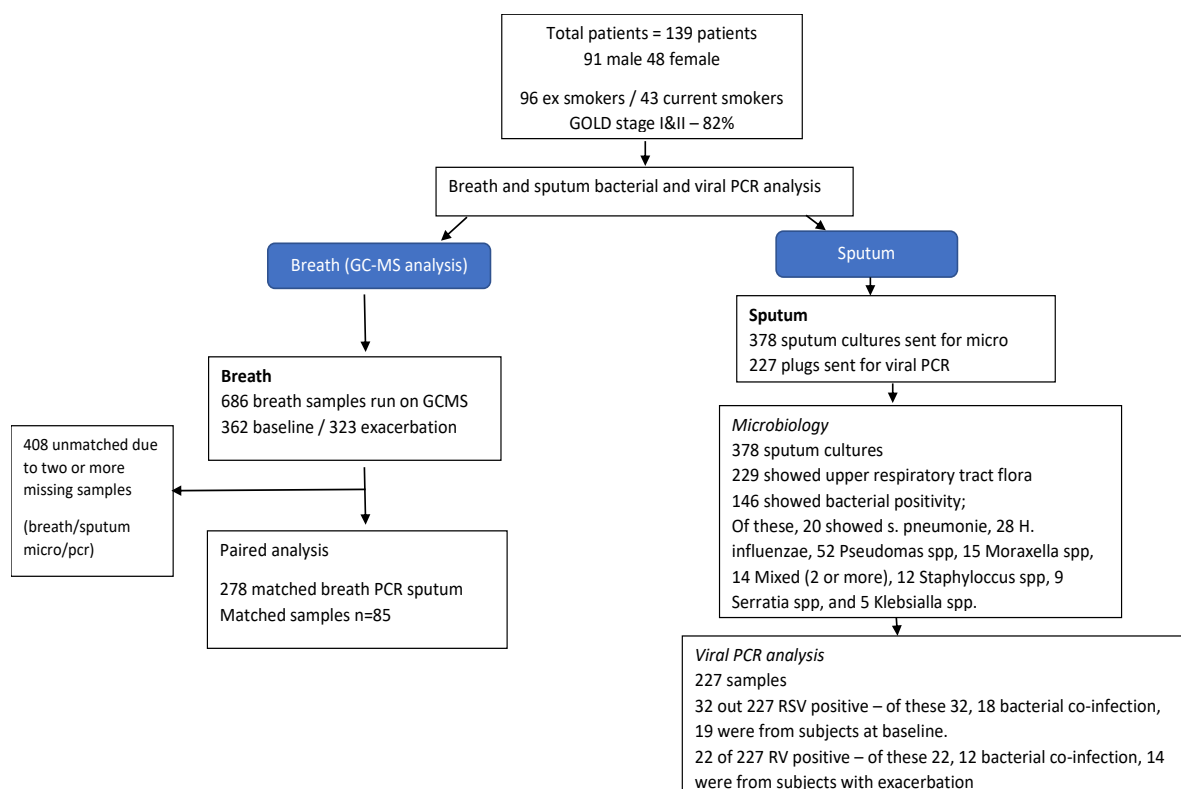


Figure 43 Overview of samples collected and analysed from cohort in relation to thesis

6.3.3 Analysis of VOCs during exacerbation compared to baseline

When comparing all subjects through clinical information given, subjects were split into two groups, baseline (healthy visits) or exacerbation (subjects called or seen at clinic and found to be unwell). Day of visit when presenting exacerbation was defined as Day 0, and further exacerbation follow ups were done at 7 and 14 days. Principal Component analysis was undertaken showing minimal separation but some discriminatory features were identified once corrected for FDR (false discovery rates) with q values <0.05.

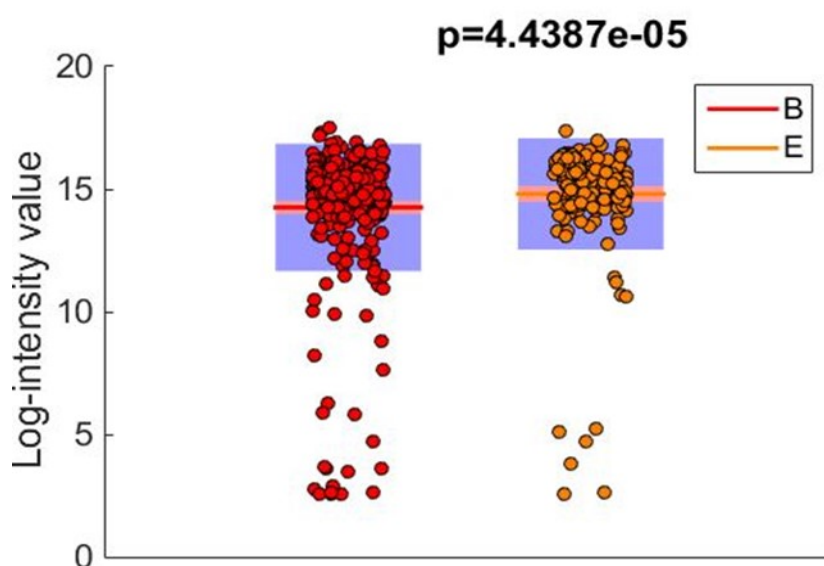


Figure 44: Induction of 2,9-dimethyl-undecane in exacerbation breath samples. Significantly higher in exacerbation breath samples (E) comparing baseline (B) samples. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

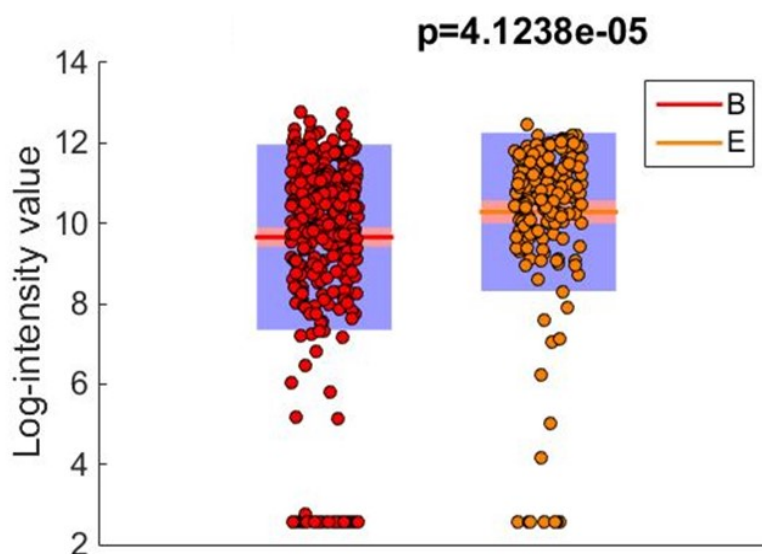


Figure 45: Induction of 1,4 pentadiene in exacerbation breath samples. Significantly higher in exacerbation breath samples (E) comparing baseline (B) samples. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

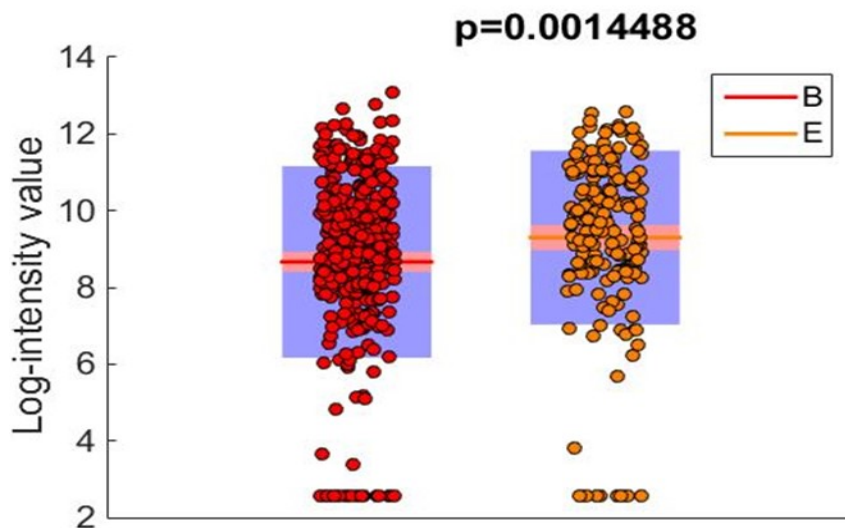


Figure 46: Induction of cyclopentanone in exacerbation breath samples. Significantly higher in exacerbation breath samples (E) comparing baseline (B) samples. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

Three examples are shown in Figures 44, 45 and 46. Over 30 VOC metabolites were identified showing significant differences between baseline and exacerbation state, with all 30 increased during exacerbation versus baseline. 10 of the 30 VOCs were able to be tentatively matched using NIST GC-MS library Table 11 shows the 10 VOCs identified as FDR adjusted and significant compounds tentatively identified. The complete 30 VOCs are described in Appendix 8.4.

Table 11: *VOCs significantly changed during exacerbation comparing baseline breath samples.*

peak_id	p value	q value	logFC (B-E)	Retention time	m/z base peak	Compound*
128	0.000407	0.031733	-0.68987	7.70	41	2-ethyl-hexanal
150	4.44E-05	0.029451	-0.54597	8.42	67	1,4 Pentadiene
464	2.75E-04	0.033126	-0.70058	18.40	85	Fufuryl alcohol, tetrahydro-5methyl-cis-
597	0.001449	0.000001	-0.63345	22.80	55	Cyclopentanone
864	0.001146	0.047525	-0.39011	33.37	83	Ethyl cyclohexane
957	5.41E-04	0.035889	-1.10379	35.43	95	Trans-9-Methyldecalin
986	0.000195	0.036903	-0.81992	36.33	97	1-methyl-2-Pentyl-cyclohexane
1040	0.000242	0.032098	-0.6745	38.28	149	Diethyl Phthalate
1041	4.12E-05	0.000001	-0.6276	38.37	57	2,9-dimethyl-undecane
1298	0.000682	0.039368	-0.63204	46.95	191	2,5-di-tert-butyl phenol

Potential identification of compounds was done through NIST (forward and reverse matching >750 and analysis of mass spectral data M/z base peak).

6.3.4 Analysis of VOCs during exacerbation, baseline and recovery

Using MATLAB and subsequent Kruskal Wallis analysis, the above VOCs were shown to have significant differences between baseline, exacerbation and recovery. Box plots below in Figures 47 and 48 show examples of 2 VOCs rising from baseline levels with exacerbation and subsequently reducing to baseline levels with recovery.

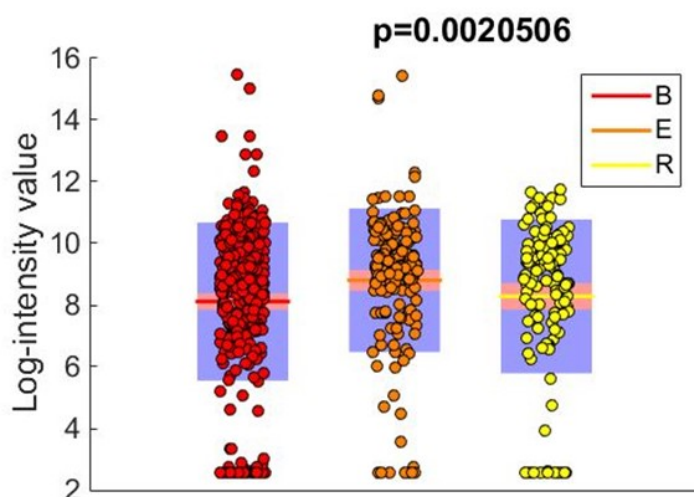


Figure 47: VOC variability overtime. There was a rise in 2 ethyl hexane (peak number 128, RT 7.70, M/Z base peak 41) from baseline through acute exacerbation and then decreasing at recovery (B= baseline breath sample, E= acute exacerbation visit from 0-1 week post symptoms, R= > 2 weeks post exacerbation breath samples. Data analysed by Kruskal Wallis test showing statistical significance between exacerbation and baseline groups with $P<0.05$. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

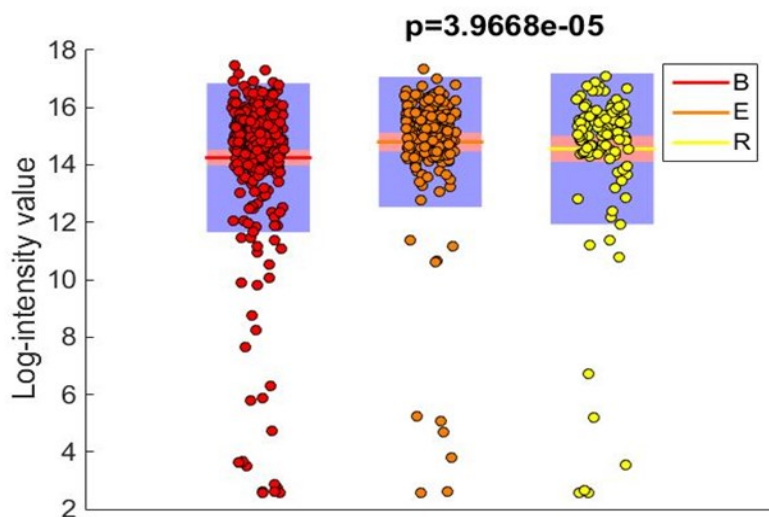


Figure 48: VOC variability overtime. There was a rise in 1,4 pentadiene (peak number 128, RT 7.70, M/Z base peak 41) from baseline through acute exacerbation and then decreasing at recovery (B= baseline breath sample, E= acute exacerbation visit from 0-1 week post symptoms, R= > 2 weeks post exacerbation breath samples. Data analysed by Kruskal Wallis test showing statistical significance between exacerbation and baseline groups with $P<0.05$.

Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

6.3.5 Temporal analysis of VOCs

Analysis was then undertaken to show the breakdown of VOC levels during the recovery weeks (from week 2 to 6 weeks post exacerbation). Below are two VOC examples ethyl hexane and 1,4 pentadiene. The changing levels these VOC levels from baseline through to recovery are graphically demonstrated.

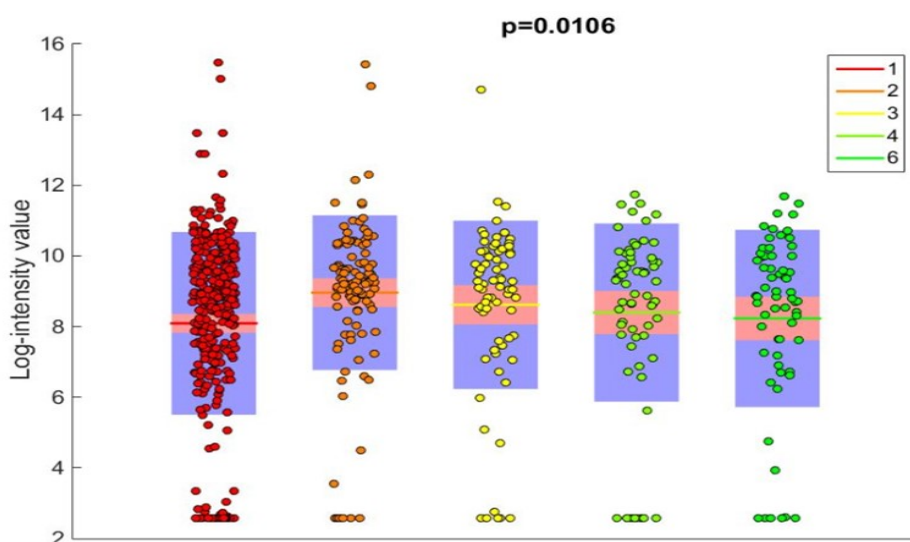


Figure 49: VOC variability overtime. There was a rise in 2 ethyl hexane (peak number 128, RT 7.70, M/Z base peak 41) from baseline through acute exacerbation and then decreasing at recovery (1= baseline breath sample, 2= acute exacerbation visit, 3= 1 week post exacerbation visit symptoms, 4=2weeks post exacerbation breath samples and 6= 6 weeks post exacerbation). Data analysed by Kruskal Wallis test showing statistical significance between exacerbation and baseline groups with $P < 0.05$. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

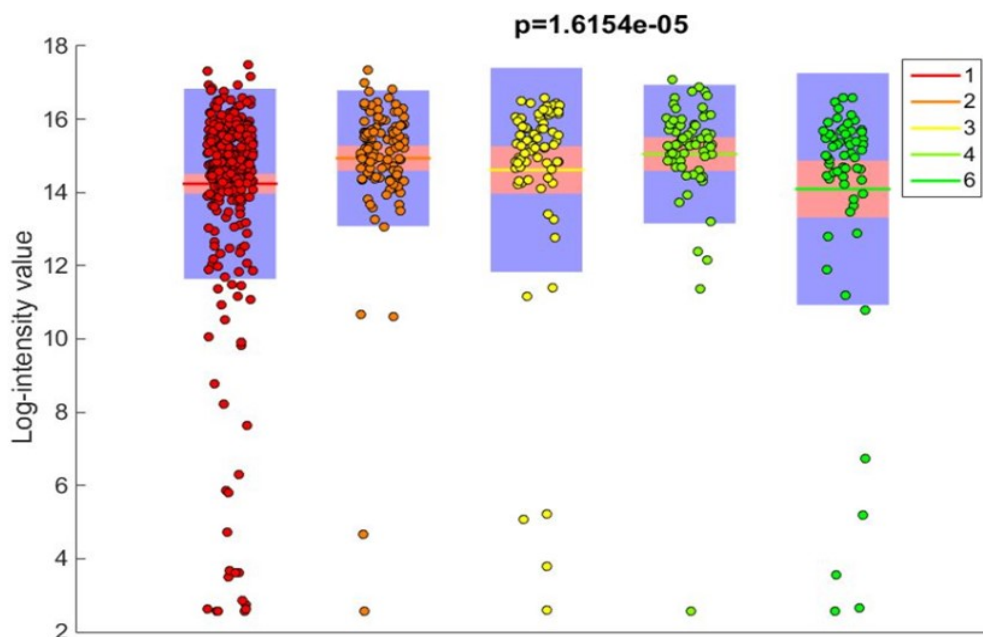


Figure 50: VOC variability overtime. There was a rise in 1,4 pentadiene (peak number 150, RT 8.42, M/Z base peak 67) from baseline through acute exacerbation and then decreasing at recovery (1= baseline breath sample, 2= acute exacerbation visit, 3= 1 week post exacerbation visit symptoms, 4=2weeks post exacerbation breath samples and 6= 6 weeks post exacerbation). Data analysed by Kruskal Wallis test showing statistical significance between exacerbation and baseline groups with $P < 0.05$ only. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

6.3.6 Analysis of VOCs in Rhinovirus positive subjects compared to rhinovirus negative subjects (irrespective of clinical stability)

When comparing breath signatures from subjects with RV positive sputum against those who are RV negative (PCR analysis of sputum irrespective of their clinical state, a stable state or during exacerbation), subtle changes in VOCs were noted but poor classification with AUC 0.62 shown. Over 40 VOCs were identified as shown below (positive p values < 0.05 and when FDR corrected, significant q values < 0.05). These VOCs increased in RV positive samples showing a difference between the two groups to allow for discrimination (see table 12).

Of interest long chain alkane VOCs shown in Table 12 which increased in RV positive samples, were also seen in Chapter 5, concerning the experimental RV studies in healthy patients. Examples of these long chain alkanes include 6,6 dimethyl undecane,

dodecane, 3,5-dimethyl undecane and are highlighted in yellow. Decane (long chain alkane) was noted to increase *in vitro* RV infection of BEAS2B cells.

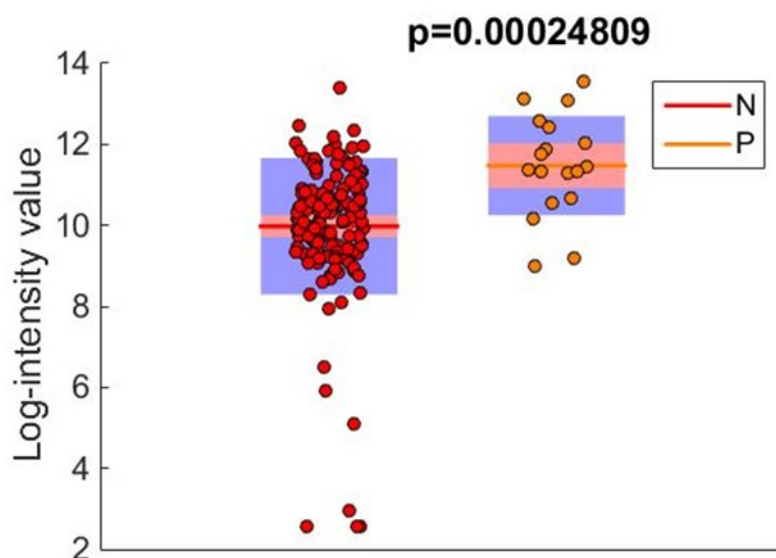


Figure 51: VOC induction in RV infection. An increase in metabolite identified at peak 667, retention time 25.73, likely phenylthene in RV viral infection (P) comparing RV negative samples (N). Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

Table 12: VOCs changes in RV positive breath comparing RV negative breath

Peak	p value	q value	Log change (N-P)	Retention time	M/z base peak	*compound
54	0.001075	0.027444	-1.88804	5.54	43	1,1,1-trifluoro-2-propanone
150	0.00298	0.036281	-0.95402	8.42	67	1,4-Pentadiene
186	0.00075	0.029287	-0.72618	9.58	76	Carbon disulphide
194	0.00426	0.040377	-1.7519	9.78	43	Cyclobutylamine
250	0.002961	0.037422	-2.96625	11.64	41	Methacrolein
370	0.005315	0.044083	-2.06657	15.56	60	Acetic acid
380	0.000206	0.030357	-1.93028	15.89	70	2 Butenal
432	0.002205	0.03216	-0.72504	17.31	44	Pentanal
498	0.003996	0.03987	-1.36849	19.45	45	Acetoin
511	0.001223	0.027974	-0.81625	19.90	70	Trans-1-Butyl-2methylcyclopropane
514	0.002061	0.031429	-0.60618	20.12	85	Octane
558	0.00106	0.027579	-0.84218	21.67	166	Tetrachloroethylene
581	0.000898	0.030569	-1.90681	22.24	45	1.2-Propanediol
612	0.002166	0.032295	-0.9313	23.54	51	Branched hexane
617	0.005317	0.043823	-1.92874	23.70	60	Butanoic acid

649	0.00232	0.033466	-0.65717	24.79	57	Nonane
659	0.004399	0.040536	-0.99744	25.38	82	3 methyl-furan
667	3.53E-05	0.046817	-1.4985	25.73	102	Phenylethyne
683	1.31E-03	0.027633	-1.07968	26.35	72	3-methyl-1acetyl-cyclopentanol
684	0.001433	0.028816	-0.82148	26.35	57	6,6-dimethyl undecane
687	0.002132	0.032147	-1.02512	26.50	58	2-heptanone
694	0.001433	0.028386	-0.8649	26.85	70	Heptanal
746	0.001537	0.02956	-0.79076	28.59	91	Propyl-benzene
789	0.00019	0.031598	-2.55842	30.33	105	1,2,4-trimethyl-benzene
791	0.000659	0.030167	-0.7816	30.54	106	Benzaldehyde
805	0.000659	0.029161	-0.82987	31.12	68	D-limonene
827	7.37E-05	0.032612	-1.02245	32.03	103	Benzonitrile
828	5.07E-03	0.042874	-0.69162	32.10	57	2-ethyl-1-hexanol
928	0.002743	0.036402	-0.69729	34.64	57	Nonanal
937	0.002653	0.036291	-1.13785	34.93	57	2-methyl-Undecane
977	0.004915	0.042354	-0.61843	36.11	57	Dodecane
1001	0.005748	0.045134	-0.76218	36.83	71	2,3,5-trimethyl-Decane
1029	0.002883	0.037144	-1.21903	37.71	57	3,5-dimethyl-Undecane
1034	0.003343	0.037282	-0.53958	37.96	57	Decanal
1069	0.004838	0.042802	-0.748	39.22	57	Tridecane
1228	0.000381	0.02406	-0.82251	44.27	104	Phthalic anhydride
1236	0.001721	0.030043	-0.8108	44.49	91	1-chloro-Decane
1250	0.000373	0.02478	-0.20968	44.97	55	1-Dodecanol

*Compounds potential identification through NIST (forward and reverse matching >750 and analysis of mass spectral data M/z base peak)

VOCs identified in Chapter 5 are highlighted in yellow

6.3.7 Analysis of VOCs during RV exacerbation (RV positive, bacterial negative) compared to baseline (virus and bacterial negative) breath

Breath samples were compared between two groups using 'viral exacerbation', clinical history as well as microbiology and viral PCR analysis showing viral positivity alone and baseline breaths samples, clinically baseline with negative sputum microbiology and viral PCR confirmed analysis. Principal component analysis was undertaken using the significant featured VOCs. The groups were misclassified slightly but AUC 0.75 was noted on ROC.

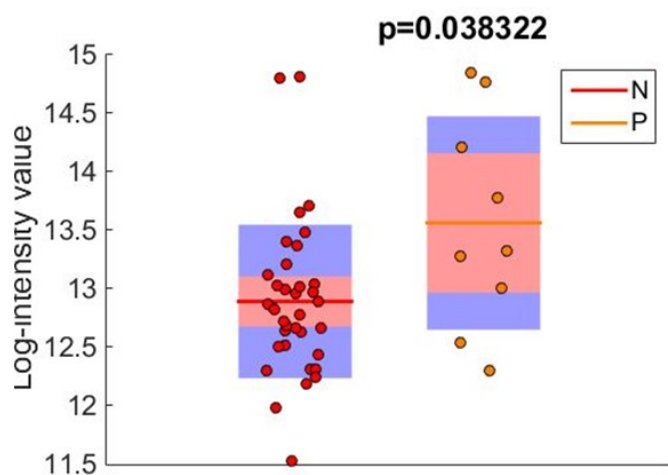


Figure 52: VOC induction in RV infection. An increase in metabolite identified at peak 720, retention time 27.71 in RV viral infection (P) comparing RV negative samples (N). Data analysed by Mann Whitney U test but not identifiable through NIST. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

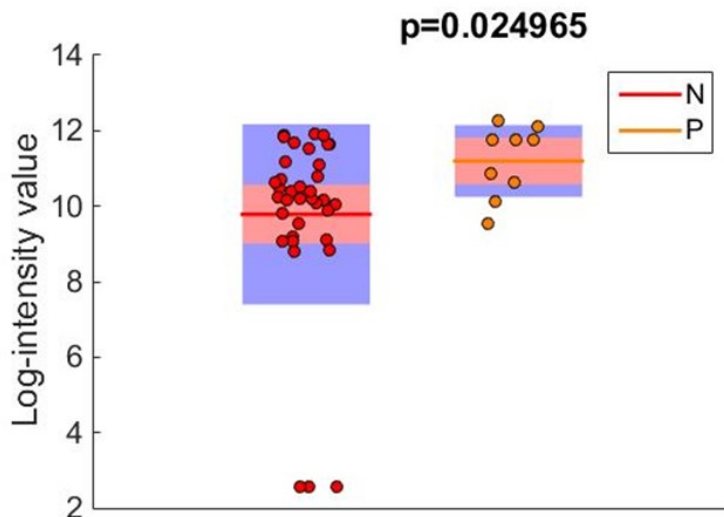


Figure 53: VOC induction in RV infection. An increase in 2,6,10-trimethyl-Dodecane increased in RV viral infection (P) comparing RV negative samples (N), a decane derivative alkane. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

No FDR q value adjusted significant compounds were identified. Statistically significant compounds were identified showing a trending towards difference between groups as

evidenced by the examples shown. Four of the 29 statistically significant compounds were able to be tentatively identified using the NIST library, shown in table 13. The full set of all 29 VOCs are shown in Appendix 8.4

Table 13: VOCs changes comparing RV positive breath samples versus baseline breath samples (negative bacteria and virus)

Peak	p value	FDr	Log fold C	Retention time	m/z base peak	Compound*
96	0.038892	1	1.23510 1	6.85	27	Hydrogen cyanide (PM 27)
601	0.0185	1	-1.20925	22.97	43	2,4-dimethyl- Heptane
680	0.041051	1	-0.53672	26.23	103	1,3,5,7-Cyclooctatetraene
810	0.038322	1	-0.67092	31.29	57	2,6,10-trimethyl-Dodecane

*Compounds potential identification through NIST (forward and reverse matching. >750 and analysis of mass spectral data M/z base peak).

6.3.8 Analysis of VOCs in RV positive exacerbation (bacterial negative) compared to RV negative exacerbation (bacterial negative)

Exacerbation breath samples, bacteria culture negative but either RV positive or negative, were analysed to see if any significant VOC metabolites visible between the above groups. 77 significant potential VOC markers were identified.

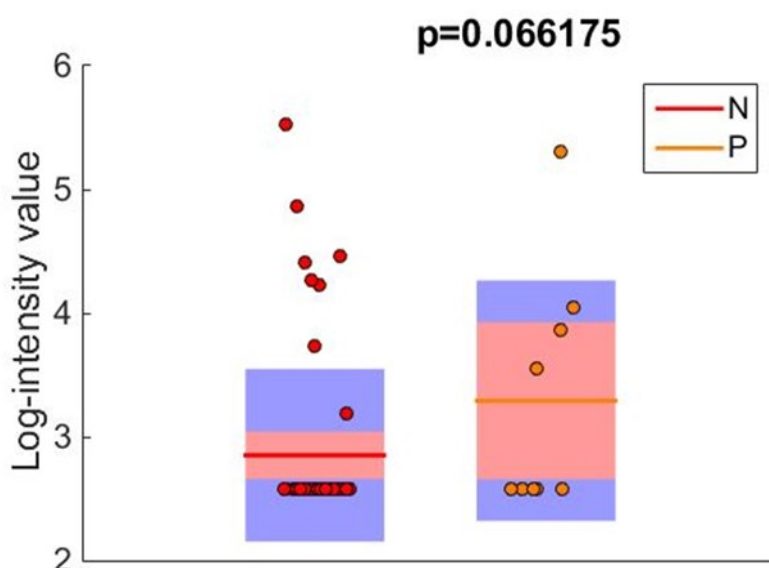


Figure 54: VOC induction in RV infection. Peak 751, rt 29.00, likely decane but did not show significant elevation in RV infection exacerbation COPD patients. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

Table 14 shows a comparison between RV positive vs RV negative exacerbation samples. When FDR corrected, none of them met the strict criteria for statistical significance (Q value <0.05) but when reviewing the data and raw plots, there is a trend towards significance. 13 of the 77 VOC peaks were identifiable when matching with NIST library the others unfortunately were a poor match when reviewing the mass spectra.

Table 14: VOCs changes comparing RV positive exacerbation breath samples versus RV negative exacerbation samples

Peak number	p value	Q value	Log change (N-P)	Retention time	M/z base peak	Compound*
56	0.044572	1	-0.10850539	5.54	60	Methyl formate
				6.38		1-en-2-ylether fluoromethyl pentafluoroprop
84	0.003619	1	-1.51457938		128	
344	0.026108	1	2.411487777	14.77	43	2 methyl 1 propanol
				15.95		Trifluoro-anhydride Acetic acid
385	0.019771	1	-2.37507975		69	
				20.17		1-(2-furanyl)- 3-Butene-1. Diol
516	0.026046	1	-0.82777771		97	
566	0.039348	1	-0.87397436	21.79	97	1-isocyano-Napthalene
601	0.002044	1	-2.55611946	22.97	43	2,4-dimethyl-Heptane
607	0.032138	1	-0.78505627	23.30	111	1-(2-thienyl)-1-Pentanone
				24.13		Acetic acid, 1- [(1dimethylethyl)thio]
633	0.049372	1	-0.81431821		57	
719	0.006481	1	-1.2746258	27.71	55	Cyclohexanone
				34.13		Phosphorous pentafluoride
906	0.015844	1	0.557125714		107	
972	0.046112	1	0.204942151	35.91	149	Ethaneperoxoic acid
				36.76		2-(methylamino)ethyl- Phenol
1000	0.022451	1	0.796198455		44	
				46.41		Hept-4yl isobutyl ester phthalic acid
1281	0.033829	1	-2.53461952		149	

*Compounds potential identification through NIST (forward and reverse matching. >750 and analysis of mass spectral data M/z base peak).

6.3.9 Analysis of VOCs in RSV positive subjects compared to RSV negative (all clinical states)

When comparing VOCs from subjects found to be RSV PCR positive (n=32) versus RSV negative (n=38) irrespective of clinical state, tentative VOCs were identified using NIST criteria. These changes were noted when comparing RSV positive subjects'

breath versus RSV negative sputum PCR subjects' breath. 38 VOCs were shown to be trending towards a difference with p values less than 0.05, but these were not significant statistically when FDR corrected (no q values < 0.05). Only 2 of the 38 were able to label tentatively with NIST library and shown in table 15. Figures 55 and 56 below show 2 VOCs, which were increased in RSV infection.

Table 15: VOCs changes comparing RSV positive breath samples versus RV negative samples

Peak	p value	q value	log correction (N-P)	Retention time	M/z base peak	Compound*
601	0.026038	0.987199	1.33702	22.97	43	2,4-dimethyl-Heptane
747	0.005481	0.909096	0.801017	28.59	149	Diethyl phthalate

*Compounds potential identification through NIST (forward and reverse matching. >750 and analysis of mass spectral data M/z base peak).

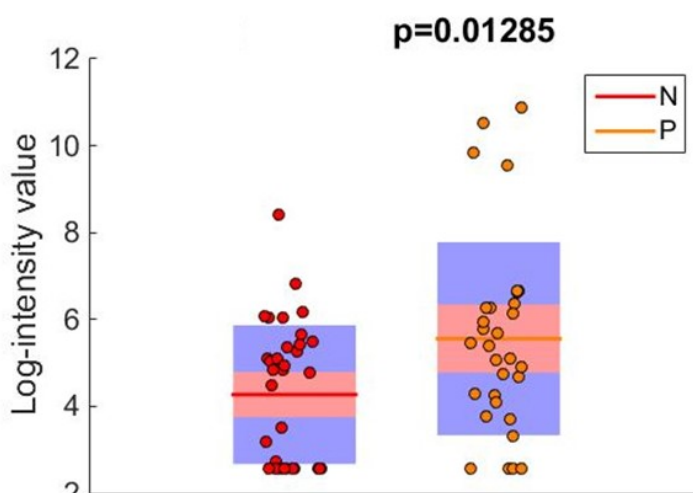


Figure 55: VOC induction in breath samples of subjects, RSV positive. Peak 591, rt 22.57 m/z base peak 45. N=RSV negative P=RSV positive. When FDR corrected q value not significantly elevated in RSV infection exacerbation of COPD subjects. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

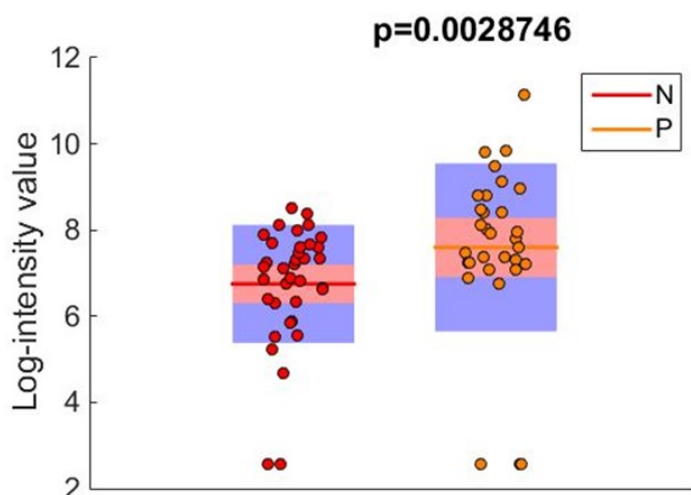


Figure 56: VOC induction in breath samples of subjects, RSV positive, peak 322, rt 13.94 m/z base peak 42. N=RSV negative P=RSV positive. When FDR corrected q value not significantly elevated in RSV infection exacerbation of COPD subjects. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

6.3.10 VOC Analysis during Bacterial colonisation: Baseline positive subjects compared to baseline negative subjects

Baseline breath samples from subjects that had positive cultures for bacteria in sputum and were clinically stable, were compared to those at baseline and grown upper respiratory tract flora in their sputum. 36 VOCs were tentatively identified using NIST criteria shown in table 16, which differ between those colonised with bacteria at baseline and those without evidence of colonisation. These compounds displayed p value significance but when FDR corrected, none were deemed to be statistically significant. There have been compounds identified in the culture work as well as well now in breath samples from those colonized with bacteria. Alcohols such as isopropyl alcohol and 2 methyl propanol were noted to increase in both. In addition, methacrolein, acetoin and aldehydes such as hexanal seen in higher levels both in vitro and human breath bacterial studies.

Table 16: VOCs changes comparing bacterial positive breath samples at baseline versus URTF at baseline samples.

Peak number	p value	Q value	Log correction	Retention time	M/z peak base	Compound
20	0.028391	1	1.734642	4.39	44	Monoammonium salt carbamic acid
63	0.014089	1	-0.94413	5.71	43	Acetone
184	0.049813	1	0.622514	9.31	45	Isopropyl Alcohol
234	0.016309	1	1.111628	11.21	56	n-Hexane
235	0.018182	1	1.043471	11.21	57	Sulfurous acid, butyl isobutyl ester
250	0.020678	1	4.890652	11.64	41	Methacrolein
288	0.045328	1	0.637264	13.08	43	2-Butanone
387	0.009502	1	-0.73652	15.95	43	Bis(1-methylethyl)ester Propanedioic acid
428	0.039134	1	1.673657	17.05	43	3,3-Dimethyl-2,4-pentane dione
433	0.006485	1	0.902781	17.38	69	Methyl methacrylate
435	0.013065	1	1.630544	17.49	43	n-Propyl acetate
454	0.049563	1	1.741336	18.15	31	2-Chloroethanol
498	0.030639	1	1.541502	19.45	45	Acetoin
511	0.010406	1	0.914283	19.90	70	Trans-1-Butyl-2methylcyclopropane
567	0.049813	1	0.645631	21.89	43	2-Hexanone
577	0.003948	1	0.744114	22.20	56	Hexanal
597	0.027674	1	0.725837	22.80	55	cyclopentanone
659	0.018182	1	1.221605	25.38	82	3-methyl-Furan
687	0.013065	1	1.511403	26.50	58	2-Heptanone
694	0.027682	1	0.530922	26.85	70	Heptanal
761	0.045328	1	-0.69771	29.29	93	B-Phellandrene
776	0.010359	1	-1.21764	29.70	43	2-methyl-propanol
800	0.010406	1	0.901599	30.97	56	Octanal
805	0.037373	1	0.482622	31.12	68	D-Limonene
862	0.024975	1	1.770217	33.19	94	Phenol
928	0.037373	1	0.654327	34.64	57	Nonanal
935	0.037373	1	0.543546	34.79	170	4-methyl-Undecane
1029	0.027682	1	0.991581	37.71	57	3,5-dimethyl Undecane
1040	0.005741	1	2.998916	38.28	149	Diethyl Phthalate
1066	0.037373	1	0.735959	39.08	59	1-(2-butyoxy-1methylethoxy) - Propanol
1069	0.037373	1	0.611507	39.22	57	Tridecane
1075	0.013065	1	1.377101	39.46	71	3-methyl-5-propyl-Nonane
1281	0.041188	1	2.447917	46.41	149	Hept-4-yl isobutyl ester

Highlighted compounds from table above seen in collection of VOCs from bacterial culture and human breath from subjects found to be colonised with bacteria.

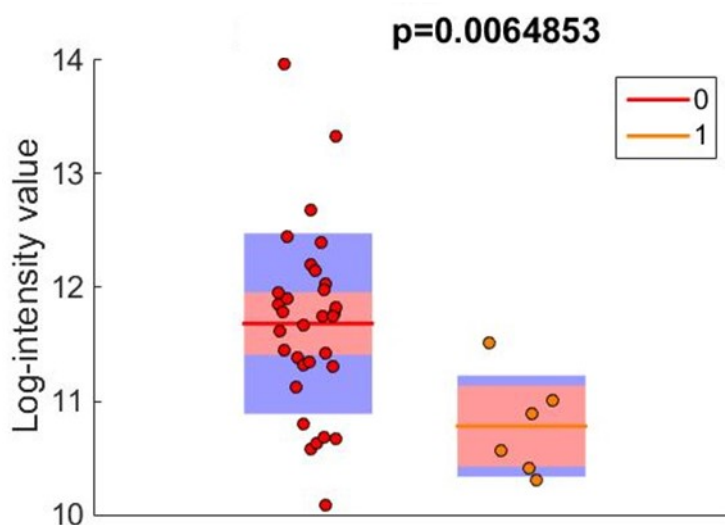


Figure 57: VOC detection in breath samples of patients colonised with bacteria. Peak 433 shows methyl methacrylate levels in both groups and the clear trending difference between the groups despite no statistical significance. 0= URTF 1= bacteria at baseline. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

6.3.11 Analysis of VOCs during Bacterial positive exacerbations (virus negative) vs Baseline (negative virus and bacteria)

Breath from the patient exacerbation visits with positive bacterial culture through sputum results (n=28) was compared to breath samples from subjects at baseline that were bacterial and virus sputum negative (n= 30). Comparison between groups with significant metabolites showed good separation between groups with ROC AUC 0.73 (see Figure 58). ROC Curve was constructed based on cross validated class probabilities of VOC sample profiles. The class probability was calculated based on multivariate pattern of discriminating metabolites to predict the probability of a sample belonging to a group of interest and the confusion matrix with class specific classification accuracies (i.e. the percentage of correctly classified samples into their respective groups) was calculated based on the probability threshold, automatically derived by a logistic regression classifier.

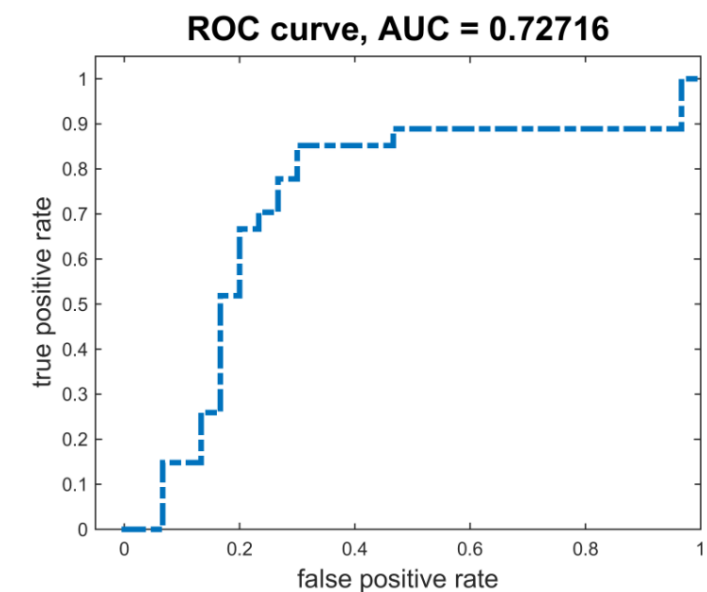
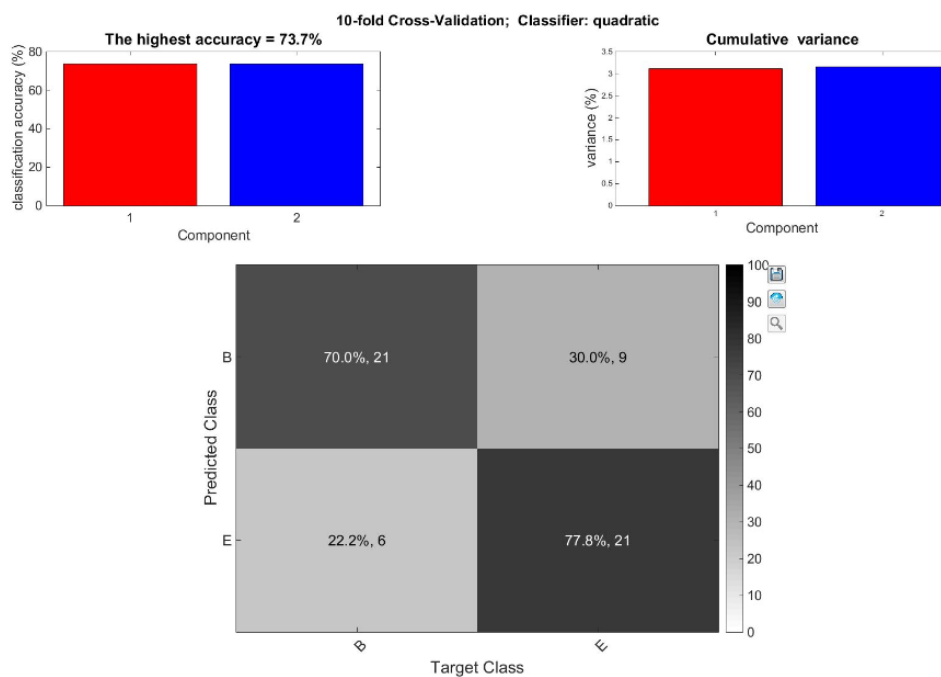


Figure 58: ROC AUC curve using VOC profile to separate bacterial exacerbation in COPD subjects breath samples compared to COPD subjects at baseline.

Confusion matrix below shows good separation and classification into correct groups baseline and exacerbations



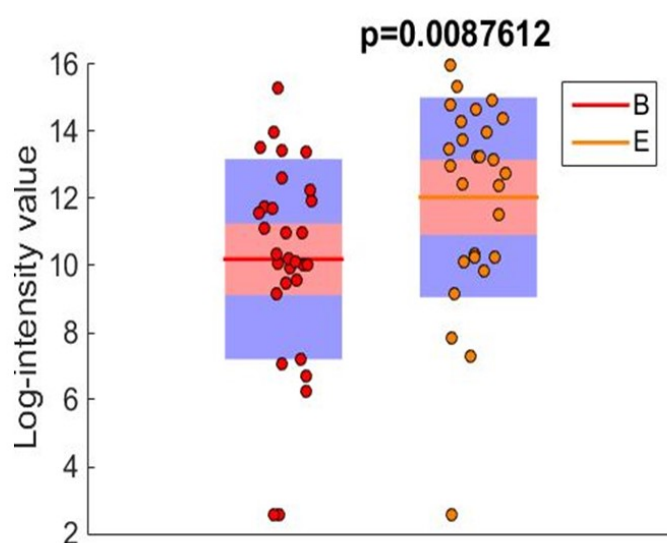


Figure 59: VOC induction in bacterial infection. Increase in acetic acid (peak 370) in acute bacterial exacerbation breath samples (E) comparing baseline (B). Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

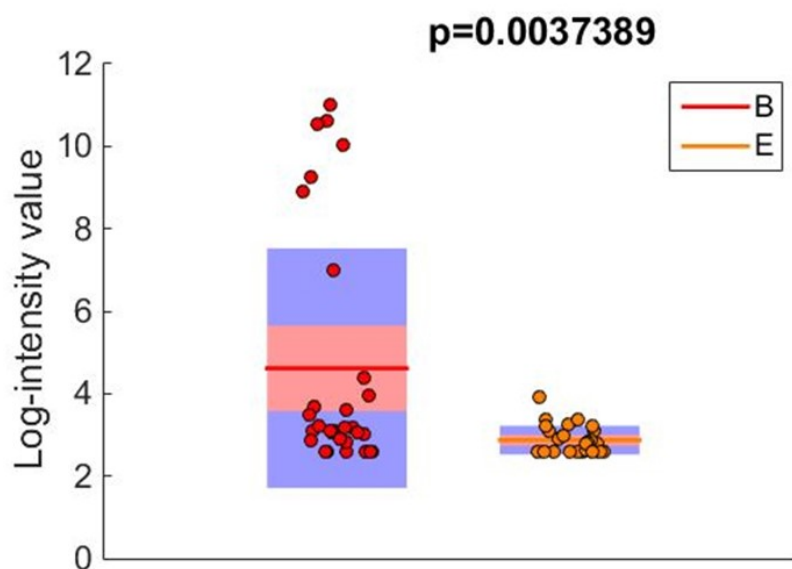


Figure 60: VOC detection in subjects with bacteria positive sputum. Decrease in propanedioic acid (peak 387) in acute exacerbation sample comparing baseline breath. Data analysed by Mann Whitney U test. Median represented by the line, 95% confidence interval for median and confidence interval for data shown by pink and blue areas respectively.

29 VOCs shown in table 17, showed significant differences with p values < 0.05. Tentatively identified through NIST, these VOCs could propose difference between bacterial exacerbations and baseline breath samples that were taken from subjects who were negative for bacteria and viruses in their phlegm at baseline visits. Several of these were seen in the in vitro agar plate bacteria VOC collection and highlighted in the table below showing potential translational role of VOC detection from in vitro to human breath.

Table 17: VOCs changes comparing bacterial positive breath samples during exacerbation versus baseline (negative bacteria and virus) breath samples.

Peak number	p value	Q value	Log change (B-E)	Retention time	M/z base peak	Compound
31	0.004226	1	-0.324384953	4.92	58	Acetone
42	0.003911	1	-1.966161017	5.22	27	Tetra-t-butylethane
47	0.012659	1	-0.872286124	5.32	43	Cyclobutylamine
58	0.039237	1	-0.647420428	5.60	41	1-butene
64	0.004444	1	-0.943975309	5.76	43	Dimethyl-Diazene
67	0.016516	1	-1.062846708	5.91	51	Pentafluoroethane
73	0.030317	1	-0.955843764	6.15	44	Hydroxyurea
114	0.044933	1	-0.905467646	7.37	101	Trichloromonofluoromethane
150	0.042384	1	-0.574950977	8.42	67	1,4-Pentadiene
169	0.045742	1	-1.443994588	8.85	58	Propanal
208	0.003721	1	-2.512424783	10.16	84	2-butyl-1methyl-pyrrolidine
211	0.002503	1	-1.763714044	10.23	28	Hydrogen cyanide
219	0.01867	1	-0.891214837	10.68	57	Butyl isobutyl ester sulphurous acid
234	0.037739	1	-0.848099989	11.21	56	n-hexane
362	0.018767	1	-0.80383105	15.27	181	Diglycolic acid, di(pentafluorobenzyl) ester
370	0.008761	1	-1.847959715	15.56	60	Acetic acid
375	0.00066	1	-1.175925134	15.65	43	Acetic anhydride
387	0.003739	1	1.745927532	15.95	43	Propanedioic acid
391	0.01742	1	-0.773663283	16.12	43	Diacetyl sulphide
648	0.016514	1	-1.030149731	24.69	55	Ethenyl ester 2-propenoic acid
747	0.036037	1	-0.791641621	28.59	149	Diethyl Phthalate
1281	0.039237	1	-1.634802668	46.41	149	Hept-4yl isobutyl ester phthalic acid

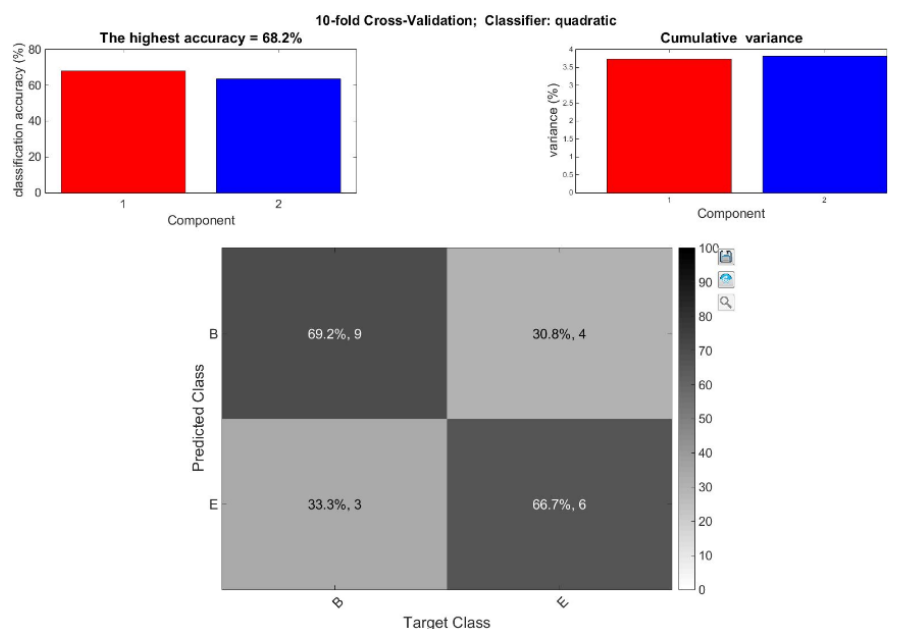
*Compounds potential identification through NIST (forward and reverse matching. >750 and analysis of mass spectral data M/z base peak).

Highlighted in yellow, compounds found in vitro and in breath of patient with bacterial positive sputum.

6.3.12 Analysis of VOCs in *Streptococcus pneumoniae* and *Haemophilus influenza* positive samples at exacerbation compared to baseline (bacteria and virus negative)

The next analysis undertaken was to examine exhaled breath from patients, who grew *H. influenzae* and *S. pneumoniae* in sputum given during exacerbation (n=8), and compare this to those subject's breath at baseline (n=12) that were proven via sputum to have negative virus and bacteria in sputum. The VOCs were not as well separated or classified when comparing all bacterial positive exacerbations with AUC of 0.62. When undertaking this subgroup analysis, no common VOCs were seen when comparing to VOCs from agar plate or *in vitro* cell infection model (table 18).

ROC curve was constructed based on cross validated class probabilities of VOC sample profiles. The class probability was calculated based on multivariate pattern of discriminating metabolites to predict the probability of a sample belonging to a group of interest and the confusion matrix with class specific classification accuracies (i.e. the percentage of correctly classified samples into their respective groups) was calculated based on the probability threshold, automatically derived by a logistic regression classifier.



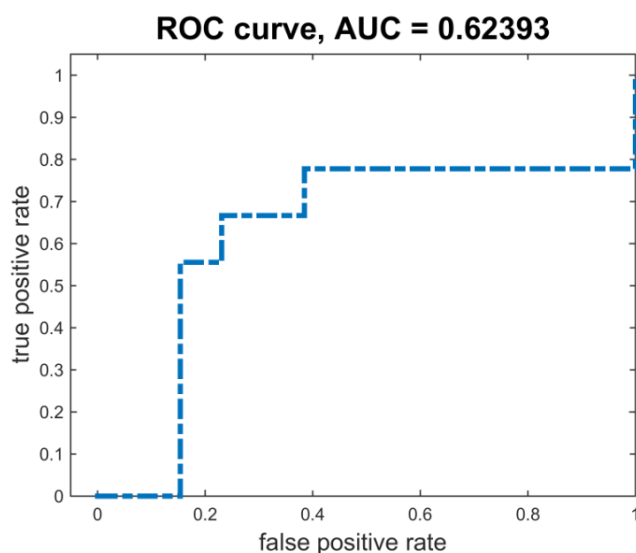


Figure 61: ROC AUC curve for *S. pneumoniae* and *H. influenzae* bacterial exacerbation of COPD subjects; breath samples compared to COPD subjects at baseline.

Table 18 shows the most significant p values for differentiating VOCs when comparing the two groups positive *S. pneumoniae* and *H. influenzae* at exacerbation versus baseline breath samples.

Table 18: VOCs changes comparing bacterial (*S. pneumoniae* and *H. influenzae* only) positive breath samples at baseline versus baseline (negative bacteria and virus) breath samples.

Peak	p value	FDR	log change	Retention times	m/z base peak	Compound*
375	0.000998	1	-1.39152	15.65	43	Acetic anhydride
1293	0.001197	1	-3.25773	46.77	221	Poor Match (MW 278)
1096	0.002329	1	-5.61379	40.30	83	Poor Match (MW 222)
195	0.002378	1	-2.43634	40.30	83	Poor Match (MW 222)
457	0.00293	1	-0.26281	18.25	69	Trifluoro-Methanesulfinyl fluoride
611	0.003579	1	-1.2083	23.49	71	Poor Match (MW 256)
31	0.003674	1	-0.56162	4.92	58	Embramine/Acetone
73	0.004527	1	-2.51716	6.15	44	Hydrogen cyanide
999	0.004539	1	-1.85782	36.76	71	Poor Match (MW 71)
315	0.006778	1	-2.42401	13.85	57	Methyl propionate
1175	0.007352	1	3.246013	42.66	85	Poor Match (MW 162)
916	0.00946	1	-1.65952	34.32	105	Poor Match (MW 244)

6.3.13 Analysis of VOCs in exacerbation and baseline samples using a Paired analysis

Paired analysis was conducted on 76 patients (samples from same subject analysed and compared to each other). Baseline versus exacerbation breath samples identified over a hundred significant (p and q values <0.05) metabolites but the top 15 most significant are shown in table 19. 1,4 Pentadiene and 1-propanol highlighted in table below have previously been seen in bacterial agar plate culture work (chapter 3).

Table 19: Top 15 VOCs changes comparing exacerbation breath samples versus baseline (negative bacteria and virus) breath samples undertaking paired analysis

Peak	p -values	q -values	retention time	M/z base peak	Compound*
12	6.91E-28	9.17E-25	3.92	29	Unknown
1034	6.37E-26	1.69E-23	37.96	57	Decanal
596	4.24E-26	1.87E-23	22.68	83	1,2,3,6-Tetrahydropyridine
1250	5.84E-26	1.94E-23	44.97	55	1-Dodecanol
184	1.06E-25	2.34E-23	9.31	45	2-Propanol
150	1.24E-25	2.36E-23	8.42	67	1,4-Pentadiene
257	4.12E-26	2.74E-23	11.9	31	1-Propanol
848	2.52E-25	4.18E-23	32.64	244	Poor match (MW 244)
928	3.21E-25	4.26E-23	34.64	57	Nonanal
1093	2.97E-25	4.37E-23	40.1	71	4,6-dimethyl-Dodecane
628	3.92E-25	4.73E-23	24.01	55	Poor match (MW 297)
773	8.30E-25	9.18E-23	29.65	85	Tetrahydropyran Z-10-dodeconoate
1144	1.39E-24	1.42E-22	41.96	57	Tetradecane
59	1.85E-24	1.76E-22	5.6	56	Poor match (MW 142)

Highlighted compounds in yellow seen in previous bacterial culture work

Long chain alkane 4,6 dimethyl-dodecane seen to increase in exacerbation breath comparing baseline breath. Long chain alkanes have been seen to increase in RV infections *in vitro*, RV human infection and in the breath of RV positive COPD patients.

6.3.14 Analysis of VOCs using regression analysis

Regression analysis was undertaken for multiple subject factors, assessing their potential effects on VOC profiles.

Subject sex was found to have no discriminatory features observed (AUC 0.50). Smoking status was analysed. Non-smokers breath and smokers breath samples were examined irrespective of clinical state and showed good separation between the groups, with VOCs giving an AUC 0.72 with ANOVA analysis. 29 FDR corrected volatiles

identified but only 12 were able to be tentatively matched (see table 20). Peak 190, 298 and 320 are the most statistically significant, smaller than 0×10^{18} (q and p values) but were not able to be matched through NIST, however gave the greatest statistical significance when analysis was undertaken.

Table 20: *Smoking related VOCs changes identified comparing smoking breath samples versus non-smoking breath samples*

Peak number	p value	Q value	Log correction (N-S)	Retention time	m/z base peak	Compound*
58	0.000544	0.036113	-0.619760029	5.60	41	1-Butene
129	0.000508	0.035503	-0.67942873	7.82	42	Methyl-Cyclobutane
150	0.000694	0.036839	-0.73065882	8.42	67	1,4 Pentadiene
189	0.00059	0.034048	-0.289012377	9.62	43	N,N,O-Triacetylhydroxylamine
199	8.34E-04	4.10E-02	-0.629314616	9.91	84	Methylene Chloride
269	5.11E-07	7.54E-05	-0.774790924	12.28	82	2-methyl-Furan
408	9.36E-07	0.000124	-0.511339453	16.42	43	N,N,O-Triacetylhydroxylamine
414	4.44E-16	1.47E-13	-1.418501153	16.63	96	2,5-Dimethyl furan
601	6.04E-05	0.007283	-0.947134272	22.97	43	2,4-dimethyl-Heptane
1062	0.000835	0.039573	-0.472186734	38.92	175	1,3(1,1-dimethyl)-Benzene

Dimethyl furan, butane, heptane, benzene have all been reported in the literature with relations to smoking related VOCs¹⁷⁰.

Further regression analysis was undertaken to assess the effect of FEV1/GOLD stage /steroid inhaler usage/number of exacerbations in a year and BMI assessed using spearman/pearson and partial least regression (PLS) regression methods. There was no correlation seen. The metabolite levels or phenotype was independent to the listed subject variables which means that any potential differences noted in VOCs levels attributable to patient status i.e. baseline, exacerbation or recovery.

The 10 peaks which showed Q value significance when comparing exacerbation breath samples to baseline had a ROC AUC~ 0.6. Using regression analysis on the 10 identified VOCs of interest, PCA linear discriminant analysis (LDA) and linear criteria

instead of MMC (because the number of variables substantially less than the number of samples) regression analysis undertaken to correct for any patient factors that might contribute to the VOC expression and the ROC curve AUC was seen to significantly improve from AUC ~0.6 to an AUC of 0.74 with an AUC > 0.7 being deemed clinically useful. Essentially using the 10 identified VOCs, a diagnostic model could be created that could help with establishing exacerbation in the future and aid clinical judgement.

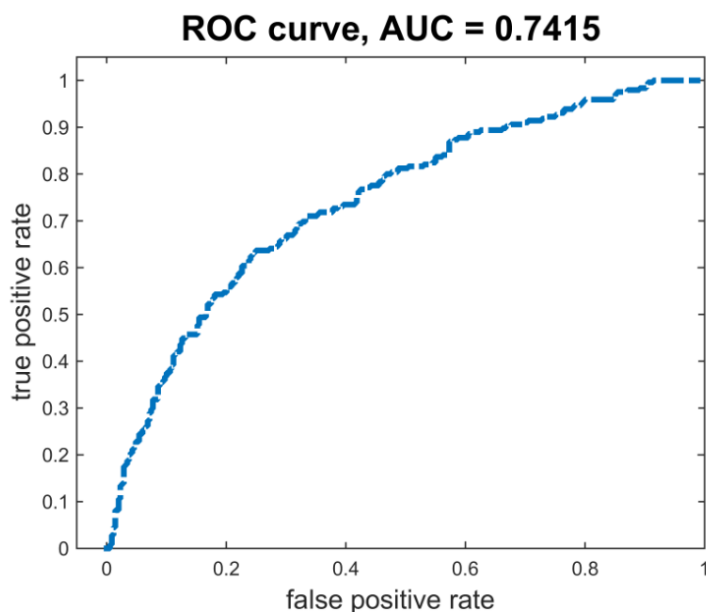


Figure 62: ROC AUC curve; VOCs from exacerbation breath samples have the diagnostic ability to be distinguish from COPD subjects at baseline (breath regressed for patient factors).

6.4 Discussion

To the best of the author's knowledge, the present study is the largest cohort following COPD exacerbations that has assessed VOCs in exhaled breath samples and the viral or bacterial aetiology of COPD exacerbations. Previous literature by Filipak *et al* suggested a set of 4 compounds: n-butane, cyclohexanone, 2-pentanone and 4-heptanone, were found to discriminate acute exacerbated COPD from both stable COPD and healthy individuals with an AUC of 0.92¹⁷¹. Additionally, this study suggested an AUC of 0.74 using a panel of 10 VOCs which also includes ketones, alkanes and aldehydes. The increase in 10 VOCs during acute exacerbation with subsequent return to baseline levels once again shows the feasibility of breath related VOCs to be used as a potential to differentiate aetiology and subsequently guide management of exacerbations. In the present study, Q value FDR corrected metabolites identified

discriminators between exacerbation and baseline breath samples as well as rhinovirus positive breath samples.

Focusing on just clinical information (excluding microbiology and virology results), exacerbation breath samples were analysed and comparing baseline samples approximately 30 VOCs were able to differentiate between the groups with FDR corrected significant statistical analysis. 10 of these VOCs were able to be tentatively labelled using the NIST library. These compounds included alkanes, alcohols, aldehydes and fatty acids groups. Sub analysis has shown the rise in these groups from baseline levels during exacerbation with subsequent fall in levels during recovery.

Breath from viral exacerbations secondary to rhinovirus, proven by sputum PCR analysis showed increase in similar chemical groups alcohols, e.g. 2 propanol and alkanes both short chain, e.g. ethane and long chain alkanes such as heptane, hexane, undecane and dodecane. These long chain alkanes were consistently seen with rhinovirus particularly *in vitro* infected bronchial epithelial cells with rhinovirus and also in clinical human rhinovirus infection (coldzyme study). A combination of alkanes were noted to increase in those COPD subjects that had rhinovirus exacerbations such as heptanes, cyclohexane and undecane. Through section six, the relationship between viral infections and long chain alkanes which are seen throughout *in vitro* and human breath observations will be explored. RSV exacerbations however, were not seen to follow the same rise in VOCs. Minimal separation between those found to have RSV positive sputum and those negative was noted. In addition, a higher percentage (19 of 32) of those found was found in baseline samples, suggesting potential presence in well individuals an almost colonised viral presence.

When evaluating the paired analysis (exacerbation vs baseline), a greater number of VOCs were noted to differentiate between the two groups. 1,4 Pentadiene was seen to be higher in both analysis of bacterial and viral exacerbation groups. This VOC was also seen to be higher in bacterial exacerbations as well as aldehydes such as propanal. Aldehydes did not feature particularly in human RV challenge or RV and RSV presence in sub analysis undertaken in the COPD cohort i.e. Virus positive subjects, their breath did not contain aldehydes through our GC-MS analysis. The sub analysis undertaken on COPD cohort for subjects with proven bacterial exacerbations followed *in vitro* results with detection of aldehydes e.g. propanal, alcohols e.g. butanol, propanol and isopropyl

alcohol and fatty acids e.g. propandioic acid. Other VOCs that have been noted to be elevated in bacterial exacerbations include sulphide compounds such as diacetyl sulphide and hydrogen cyanide which also have increased *in vitro* plates and cells infected with bacteria. Additionally, some of these VOCs were seen to increase in viral infections. RV positive patients with exacerbations were compared to those that were RV negative, and at baseline were shown to have increased levels of acetic acid, alcohols e.g. 2 methyl 2 propanol as well as hydrogen cyanide.

The long chain alkane undecane was noted to be elevated in RV confirmed COPD exacerbations consistent with *in vitro* and clinical human RV challenge results. Long chain alkanes undecane and other decane derivatives, 2 methyl-undecane and trimethyl dodecane, were increased in RV COPD infection breath samples comparing baseline breath samples and were not seen in bacterial infection. Decane and its derivatives were seen to increase in all three RV models studied, and their potential importance as diagnostic VOCs is discussed in 5.2.4.

Fatty acids (acetic acid and propanedioic acid), aldehyde (propanal) and ketones such as acetone were found to be higher in bacterial proven exacerbations (ROC AUC 0.73) and these metabolites are consistent with *in vitro* experiments which once again in more controlled environment identified highly diagnostic ROC AUCs >0.8 (limitations of cohort discussed below). These metabolites are energy substrates that have been discussed previously, and are thought to be needed for bacterial replication and propagation.

Due to the nature of the COPD cohort and study design, it was not always logistically feasible to analyse all breath samples within 24 hours, hence a degree of VOC degradation is possible. These limitations could explain the reduced AUC with the data set but the promise of breath as a non-invasive diagnostic certainly has been reaffirmed. *In vitro* data concerning VOC collection was all undertaken by myself and there was more control of instrumental analysis so it is hard to know how much of the factors below contributed to a ROC AUC of 0.8/0.9. In addition, all breath samples of the RV human cohort were collected and run on the GC-MS instrument in in a shorter time frame due to the nature of the study design which too may have contributed to the consistency of the data in the RV human data challenge. Clearly, further work is required to confirm the diagnostic importance of the VOCs identified. The p values were extremely low at times between compared groups. The p value is the function of the number of samples

and effect size (e.g. median level differences between the groups of interest). It tends to become smaller as sample size increases. With a sufficiently large number of samples tiny differences can be statistically significant. However, I accept that in the examples shown were the p values were extremely low, the effect size (although statistically significant) was rather small to be of clinical relevance.

Also of note through the analysis was the tentative identification of an increase of 'contaminant' VOCs detected by the GC-MS system. Siloxanes, silanols and silanes are found as a result of the rubber tubing used to connect TD tubes to the gastrocheck steel bags and subsequent collection occurs through the use of air flow pumps to entrain collected breath on the thermal desorption tubes. The GC-MS instrumental set up also has these known contaminants from the column. These contaminants are listed in the results table and the appendices section for completeness but are of no clinical significance. In this chapter, and the RV human challenge study, increased concentrations of long chain alkanes have been observed in viral infections despite the above limitations. These include; instrumental variances such as time drift and column change associated changes to retention times as one would expect with MS analysis, timing of sample analysis variance i.e. particularly with regards to breath collection in the human cohort, and time to analysis to avoid degradation of volatiles and the lack of GC-MS standards including a calibration curve and breath sampling standardisation i.e. subjects were sampled at any time with no restrictions on diet, smoking etc. prior to sampling. Through the nature of the COPD cohort it was never possible to collect all breath samples from subjects. The majority (at least 2/3rds were collected by myself) but the clinical assessments were not all undertaken by myself. This could have led to inter clinician variance in terms of assignment of patient status. All breath samples however were analysed by myself on the GC-MS instruments.

The importance of the regression analysis undertaken is most promising. The independence of VOC profiles and levels irrespective of patient sex, age, BMI, severity of COPD and frequency of exacerbation gives promise. The incorporation of those individual patient factors that may potentially influence VOCs, and subsequent regression analysis have shown a clinical applicability through the potential future development of diagnostic tests to aid clinical management.

6.5 Conclusion

The COPD cohort demographics represents a varied population that would be in keeping with and representative of community respiratory patients with COPD in secondary care. The majority of patients have GOLD stage II or III (82%) disease as you would expect within a population of secondary care patients under hospital follow up. The prospective longitudinal study allowed for the first time VOC analysis in a COPD population focusing on exacerbations and investigation of their potential to be used as a diagnostic tool. The heterogeneous nature of the cohort provided the opportunity of assessing the feasibility of breath testing and analysis. Through this study there has been identified the ability through groups of VOCs to differentiate between COPD subjects exacerbating comparing when well, those having a viral exacerbation and those identified having bacterial positive sputum and exacerbation episode. In addition, VOCs identified have shown consistency in groups between the various study designs suggesting huge promise for the potential implementation of breath testing in the future to aid diagnostic decision making.

7 Summary, conclusion and future direction

7.1 Summary and Conclusion

In this research, mass spectrometric interrogation of multiple infection models has been undertaken including headspace above plates culturing bacteria, headspace above infection of airway epithelial cells with common respiratory pathogens and exhaled breath from healthy subjects inoculated with RV. VOCs have also been analysed from patients recruited from the London COPD cohort at baseline and during exacerbations (bacterial or viral aetiology was identified when subject was able to expectorate through microbiology and PCR analysis on sputum). A novel and reliable method for the collection of VOCs from *in vitro* models was established. The results demonstrate VOCs from similar chemical functional groups unique to bacteria and viruses across the various experimental models. The research confirms there is significant translational potential of VOCs to be utilised for point-of-care diagnostics in acute respiratory tract infections including pneumonia and viral and bacterial exacerbations of COPD. There have been decades of technical challenges which have led to lack of progression in the area. Breath analysis has often had absence of standard analytical methods that can address the variation in results or the accurate identification and quantitation of target compounds. Some methods require sample preconcentration by adsorption (on sorbent traps and coated fibers) or by direct cryofiltration because concentration of many substances ranges from nmol/L to pmol/L. Collection and preconcentration methods for breath samples are often considered more difficult to standardise and subsequently with these challenges minimal advances have been seen in progressing breath analytics in diagnostics.

Investigation of VOCs in bacteria and viruses has been undertaken previously^{110,119,120}. However, the techniques and longitudinal approach employed in the present project are novel. Different pathogenic bacteria produce specific VOCs when cultured on agar medium. The experimental methodology including cell culture and infection of airway epithelial cells (BEAS2bs) with bacteria and viruses allowed a like-for-like metabolite comparisons. The tentative identification of VOCs requires complementary standard studies, which have been undertaken by our group for propanal, decanal, nonanal, ethanol, decane and dodecane. Identified compounds are purchased and known

concentrations are added to thermal tubes using the standardised loading methods (described in Chapter 3) and run through GC-MS using the same programme for analysis. Retention times are then compared to ensure accuracy of identification and labelling of VOCs. The *in vitro* work has also shown feasibility of employing GC-MS to differentiate between bacterial pathogens based on their VOC profiles.

The results indicate that *S. pneumoniae* and *H. influenzae* can be differentiated through their individual VOC profiles (particularly from the *in vitro* culture headspace VOC analysis); aldehydes, alcohols and ketones were prominent groups and in keeping with VOCs identified in previous studies^{119,129}. The *in vitro* studies of *S. pneumoniae* identified 30 peaks through the analysis of GC-MS of chromatograms, with 26 compounds being identifiable through NIST. *H. influenzae in vitro* agar plates identified 13 peaks on the GC-MS chromatograms. Characterization of volatile metabolites by Filipak *et al*, comprised of alcohols, aldehydes, hydrocarbons, ketones and sulphur containing compounds, demonstrated significant overlap with compounds identified in my *in vitro* work¹¹⁹. Bacterial infection models have demonstrated VOCs from aldehydes, short alkanes (C4-7), alcohols, ketones and sulphide groups consistently across all models.

Increased presence of aldehydes in bacterial infection is seen through analysis of headspace of bacterial plates as well as in the exhaled breath of COPD patients with bacterial-proven exacerbation. Filipiak *et al*. investigated VOC release from the headspace of *Staphylococcus aureus* and *Pseudomonas aeruginosa* cultures. They observed increased concentrations of acetic acid, 3-methylbutanal, 2-methylpropanal, acetaldehyde and methyl-2-butanal in the headspace of *Staph. aureus* cultures¹²⁵. Previous work by Kumar (Submitted thesis Imperial College 2014, <http://hdl.handle.net/10044/1/24837>) demonstrated significantly increased concentrations of butanal and also detected pentanal, and heptanal using SIFT-MS in patients with a diagnosis of community-acquired presumed bacterial pneumonia.

Bacterial production of aldehydes has been previously recognised. Singh *et al* have proposed the existence of a bacteriological pathway for ethanol oxidation, i.e. ethanol is oxidized by alcohol dehydrogenase (ADH) from intestinal bacteria resulting in high intracolonic levels of reactive and toxic acetaldehyde¹⁷². Their study investigated aldehyde dehydrogenase (ALDH) activity, acetaldehyde consumption and production of

acetate by aerobic bacteria (n = 27) representing the normal human colonic flora. Reactive oxygen species (ROS) are continuously generated within living systems and the inability to manage ROS load results in elevated oxidative stress and cell damage. Oxidative stress is coupled to the oxidative degradation of lipid membranes, also known as lipid peroxidation. This process generates over 200 types of aldehydes, many of which are highly reactive and toxic. Aldehyde dehydrogenases (ALDHs) metabolize endogenous and exogenous aldehydes and thereby mitigate oxidative/electrophilic stress in prokaryotic and eukaryotic organisms. The bacterial environment, and ability to metabolize, or increased concentrations of toxic carcinogenic aldehydes, have been reported in patients with lung, bowel and liver cancer within exhaled breath and blood, respectively¹⁷³. The metabolism of cancer cells is in many ways different than in healthy cells. Some metabolic pathways are intensified, and others are inhibited. Differences of ADH and ALDH activity between cancer cells and healthy tissue may be one of the factors intensifying carcinogenesis¹⁷⁴. Four aldehyde dehydrogenase (ADLH) isozymes – 3B1, 4A1, 9A1, 3A2 – were ranked in the top 100 most differentiating genes out of 25003 (n=118). Taken together, it suggests an inadequate response of these pathways in OG cancer to an in-situ interaction, such as bile-acid refluxate or attendant inflammation. Moreover, these findings are in keeping with recognised epidemiological associations between germline allelic deactivating mutations in ALDH isozymes and OG cancer in East Asian populations¹⁷⁵. The presumed role of bacterial proliferation and changes in the microbiome, and subsequent higher levels of aldehydes, has been proposed as a mechanism for pro carcinogenic pathways¹⁷⁶.

The proposed mechanism is that aldehydes produced by bacteria can undergo additional reactions with thiol amines to form toxic adducts in cellular proteins¹⁷⁷. Longer-chain aldehydes are also produced in animals and plants by various other non-radical mechanisms, including catabolism of sphingolipids via sphingosine-1-phosphate, from which the α,β -unsaturated aldehyde *trans*-2-hexadecenal is generated, and the action of myeloperoxidase and hypochlorous acid (HOCl) on plasmalogens (2-chloro-aldehydes in this instance). Sphingolipids are a structural component of cell membranes. They are similar to glycerophospholipids comprising of sphingosine which itself, consists of fatty acids, phosphates and amino alcohols, a key source of energy for the Krebs cycle¹⁷⁸.

Fatty acids are oxidized to hydroperoxy fatty acids in animal and plant tissues by several different lipoxygenases and cytochrome P450 enzymes. Polyunsaturated fatty acids can undergo autoxidation by free radical chain reaction mechanisms. Through KEGG pathway analysis (seen in appendices section 8.5) the proposed generation of aldehydes from fatty acids, the subsequent generation of alcohols through ADH and alkane production can be extrapolated. Pyruvate produced by fatty acid is a source of energy production in the Krebs cycle and previous studies have reported its crucial role in the propagation of viruses such as influenza¹⁷⁸.

Viruses are living entities and do not have their own metabolism when compared to bacteria and need a host to replicate¹⁷⁹. Viruses dramatically modify cellular metabolism in their host upon entry into cells and this may account for some of the differences observed in VOC analysis. Bacteria are able to generate their own energy substrates and short chain alkanes were observed in all bacterial models¹³⁹ compared to viruses. High levels of glycoproteins and other lipid material are required by viruses to provide the necessary energy for virion production¹⁴⁶. Spencer *et al* have previously demonstrated a virus related increase in fatty acid synthesis¹⁴⁹. Subsequent fatty acid degradation of endogenous lipids could be a potential source for the longer chain alkanes noted in RV infection. The BEAS2b RV16 infection experiments have shown that decane is induced by RV infection *in vitro* and also detectable at increased levels versus baseline following experimental RV infection in healthy subjects. Previous literature has identified the presence of decane with regards to antiviral activity¹⁴⁴. A decane derivative, Rimantadine (alpha-methyltricyclo-[3.3.1.1/3.7]decane-1-methanamine hydrochloride) may be effective for both the prevention and treatment of influenza A infection in human strains (H1N1, H2N2 and H3N2) and has been associated with significant reductions in nasal secretion viral titres, maximum temperature, time until defervescence (mean, 37h shorter), and systemic symptoms compared with placebo¹⁸⁰. There is no proposed method of action suggested.

The results of the RV16 infection in healthy volunteers, and subsequent VOC analysis of exhaled breath, has shown an upregulation in decane and decane-related VOCs during acute infection with complete resolution upon return to baseline levels. This suggests that decane may be a biologically plausible marker of the host response to viral infection. Muller Decker *et al* postulated that decane inhibits viral transcription (affecting phosphorylation pathway) through tricyclo-decane-9-yl-xanthogenate (D609)

and inhibition of vesicular stomatitis virus (VSV)¹⁴⁴. Husson *et al* reported the modulation of host defence against bacterial and viral infections by polyunsaturated fatty acids¹⁸¹. This further highlights the potential role of fatty acids in the development of infection and resultant VOC production.

The detection of decane or decane derivatives and other long chain alkanes (C10 to C14) has been seen in all three viral models in this research. Long chain alkanes, however, were not a feature of bacterial infection. This appears to be a feature of host response rather than primary pathogen VOC production.

Suggested decane production pathways through KEGG pathway analysis identified include:

Sphingosine lyase hydrolyses sphingosine-1-phosphate (S1P) to ethanolamine phosphate and 2-hexadecenal. Chemical conversions of 2-hexadecenal to decane involving hydrolysis is a potential through hypochlorous acid (although this isn't very physiological). Enzymatically there is, however, a possibility. If 2-hexadecenal, as a fatty aldehyde, is then reduced by an aldehyde oxidase, this could subsequently generate an alkane. Interestingly, certain cytochrome P450 isoforms have this enzymatic activity. In the literature, there is reference to a deformylating aldehyde oxidase, which can release hexanoate, rather than formate resulting in the production of decane. The differing interaction between bacteria and virus and host and subsequent enzymatic activity could be an explanation as to why alternative mammalian enzymes result in variable lengths of alkane chain after or before the oxidase reaction catalysing formate release.

One issue worthy of discussion is whether or not a future diagnostic test would rely on several validated VOCs or a broader collection of VOCs that suggest a common pathway or process. One example is decane, undecane and dodecane and their derivatives. Decane was initially identified as a VOC specific to viral rather than bacterial infected cells in vitro, and was initially thought to have potential as a viral specific VOC. In the study of experimental RV challenge, decane was also identified along with other derivatives including undecane, dodecane, and various substituted undecanes. Decane was not observed in RV positive COPD sputum samples during exacerbation, however several other closely related derivatives were, namely 2,6,10 trimethyl dodecane, 3,5 dimethyl undecane and 6,6 dimethyl undecane. It is possible that one VOC may be too

short lived to be useful as a diagnostic on its own, but rather the observation of several related VOCs indicative of a common pathway may be more powerful as a diagnostic tool. Decane, undecane and dodecane represent alkanes with 10, 11 and 12 carbon atoms. Complex isomers also exist, for example decane has 75 isomers. At the time of writing, it was not clear which chemical reactions operating in biological systems could explain the conversion of decane to undecane or dodecane, etc, or how various substitutions of undecane etc could be produced in the respiratory systems studied. It is reasonable to suggest however, that the observation of several decane, dodecane or undecane (or derivatives) together may strongly suggest viral, rather than bacterial infection and these VOCs should be considered as diagnostically important collectively.

The VOCs identified *in vitro* and *in vivo*, with respect to infection are, in addition, also likely investigated volatile markers of oxidative stress and subsequent epithelial cell damage, as this occurs with progression of infection. Oxidative stress is typically due to production of reactive oxygen species (ROS) such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2) by activated neutrophils responding to foreign agents, such as bacteria, resulting in cell damage¹⁸². These ROS are known to cause damage to macromolecules within cells. As well as bacterial metabolism, there is an element of host interaction between bacteria and viruses leading to the specific development of VOCs. There is, however, significant overlap in the production of compound groups of alcohols and fatty acids by bacterial and viral infections. 1,4-pentadiene was found to be in higher significant levels in both bacterial and viral exacerbation breath signatures in the COPD cohort. 1,4 pentadiene derivatives from curcumin were found to exhibit inhibitory activity against two plant viruses and its structure is being considered as a potential lead for discovery of new antiviral agents¹⁸³.

Compounds such as alcohols and fatty acids are detected consistently in both viral and bacterial infections in both *in vitro* and *in vivo* breath collection models. Alcohols, 1- alcohols, are produced through α - or β -oxidation of fatty acid derivatives through acetyl-CoA¹³⁸. Fatty acids are precursors for a variety of important building blocks including phospholipids, sphingolipids and sterols. The degradation of fatty acids yields a high amount of ATP and their detection could be as a result of their presence as an energy substrate facilitating bacteria and viruses and their interaction with the host. Saturated long chain fatty acids C26–C37 have been shown to be required to increase viral activity and infectivity of human cytomegalovirus¹⁵⁰.

The consistent findings of bacteria and virus-specific VOCs provides a real feasibility towards further refinement and development of an exhaled breath test to identify underlying aetiology in patients with a COPD exacerbation. The change in specific VOC concentrations, from baseline through exacerbation and then back to 'normal or baseline levels', provide the potential for a diagnostic test for clinical practice. There are secondary benefits of employing such a test; including the ability to rationalise antimicrobial therapies, thus reducing antimicrobial resistance, reducing drug related costs and decreasing the incidence of antimicrobial related complications. The ability to differentiate exacerbation from baseline exhaled breath samples using the 10 most significant VOCs, with a ROC AUC of 0.74 was also a significant finding. ROC AUC greater than 0.7 in most clinical settings provides fair confidence in the functional role of a diagnostic test¹⁸⁴. These VOC biomarkers would be akin to cardiac troponin and would further guide the management of these complex patients in acute primary or secondary care.

7.2 Future directions

The work undertaken in the present study has identified potential VOC biomarkers of both bacterial and viral infection, as well as potential markers of COPD exacerbation; the identified VOCs need to be further tested in validation studies. There were several limitations (previously discussed in Chapter 5) of the prospective cohort study and through this experience, there are future opportunities to refine VOC compound identification, through standardisation of breath collection and time to analysis on GC–MS. Further improvements in clinical assessment and increase in sputum sample retrieval during COPD exacerbation will also allow for more reliable assignment to correct clinical groups. This prospective follow-up study undertook an untargeted GC–MS analysis of VOCs. However, certain compounds of interest have been identified through this research and different analytical platforms, including PTR-TOF-MS and SIFT-MS could be employed for quantification.

For each potential VOC biomarker, future studies must establish:

- (i) the mechanism(s) governing their production in both healthy and disease states
- (ii) sources of methodological variability that may reflect the unique kinetics of their release from the body

- (iii) the performance of the chosen analytical platform
- (iv) the value of clinical detection for that particular disease.

The opportunity to detect VOCs within exhaled breath and other biological samples remains an attractive prospect for non-invasive disease diagnosis and monitoring. Improved mass spectrometry techniques (GC-MS-TOF) are being developed to improve sensitivity, limits of detection and subsequent identification. Through cross-validation with different analytical platforms by multiple laboratories, the likelihood of a clinically relevant VOC biomarker is significantly increased.

In addition, further studies evaluating effect of antimicrobials on VOC production may provide invaluable data for guiding judicious duration of therapy could be undertaken. Currently, there is only subjective non-evidence based advice and investigating VOC levels in relation to antimicrobials may provide useful markers to guide clinicians on initiation and duration of treatment.

VOC collection from a controlled RV infection model could also be proposed in the future. Mallia *et al* have previously undertaken a COPD RV infection challenge model³⁶. They demonstrated a controlled viral infection model with measurement of known viral replication cytokines. The ability to use proven and known sampling methodology with the ability to compare and correlate with identified VOCs over time could prove invaluable. Emergence of secondary bacterial infections in the model in around 60% of volunteers day 12-21 after initial viral infection was noted by Mallia *et al*³⁶. Collection of VOCs in a controlled RV COPD infection model may validate both viral and bacterial VOCs in the same patients in one model. The integration of symptoms, virological and bacteriological data would also be available to test whether these specific VOCs are produced in relation to each different pathogen class.

8 Appendices

8.1 Cohort demographics and Inclusion and exclusion criteria

subject	age	sex	smoke now	pack years	chronic sputum	Inh steroid dose	Steroid nasal	weight	height	fev1	Bmi	Gold
866	71.24	1	0	9.8	0	1000	0	88.75	1.595	1.62	34.88566	2
924	83.88	1	0	44	0	0	0	72	1.81	0.91	21.97735	3
933	76.98	0	1	5.5	1	1000	0	85	1.58	1.02	34.04903	2
951	73.41	1	1	118	1	1000	0	62.9	1.783	1.7	19.78554	2
2062	89.03	1	0	0.5	0	400	0	54.8	1.564	0.61	22.40305	3
2097	73.88	1	0	63	0	1000	1	62	1.61	0.81	23.91883	3
2115	67.43	1	1	25.5	1	0	1	99.55	1.785	2.17	31.24387	2
2116	59.47	1	1	150	0	800	0	70	1.8	3.32	21.60494	1
2118	72.53	1	0	2.8571 43	0	1000	0	91.55	1.627	1.66	34.58464	2
2123	79.86	1	1	13	1	1000	0	80.5	1.75	0.7	26.28572	4
2141	84.32	1	1	21	1	0	0	66.95	1.754	1.13	21.76163	3
2144	72.21	0	1	45	1	1000	0	100.5	1.54	0.93	42.37646	2
2202	76.58	0	0	72	1	1000	0	56.7	1.67	0.64	20.3306	3
2325	72.77	0	1	15.25	0	1000	1	55	1.68	1.24	19.48696	
2329	85.44	0	0	6.25	1	500	0	64.5	1.52	0.85	27.91724	2
2405	70.64	0	1	50	0	500	0	40.1	1.6	1.22	15.66406	2
2420	77.03	0	0	39	0	1000	0	58	1.585	0.65	23.0871	3
2441	66.56	0	0	45	1	1000	0	91.9	1.589	0.97	36.39718	3
2448	70.49	1	1	99.75	1	1000	0	104.3 5	1.81	1.08	31.8519	3
2456	72.85	0	0	14	1	500	0	47.4	1.57	1.14	19.22999	2
2457	76.17	0	1	18	1	0	0	90.55	1.541	1.27	38.13144	0
2467	76.45	1	1	315	1	400	0	68	1.69	0.65	23.80869	4
2470	75.8	0	0	54	0	500	0	51.9	1.55	1	21.6025	2
2479	82.71	1	0	9.5	0	1000	0	104.7	1.71	1.55	35.80589	2
2480	71.9	0	1	53	0	800	0	47.3	1.572	0.59	19.14062	3
2485	72.49	1	0	48.1	0	0	0	97.8	1.76	1.8	31.57283	2

2490	72.99	1	0	0.1	0	0	0	72.7	1.87	2.79	20.78984	0
2507	70.83	1	0	39	0	1000	0	97.9	1.72	1.58	33.09221	2
2513	68.79	1	1	21	1	500	0	107.9	1.68	1.2	38.22988	3
2517	73.92	1	0	45	0	200	0	71.25	1.65	1.21	26.1708	3
2529	60.55	1	1	34.5	1	200	0	82.4	1.76	1.06	26.60124	3
2535	70.56	0	0	84	0	400	0	113.7	1.6	1.48	44.41406	2
2536	81.03	0	0	0	0	1000	0	57.6	1.73	1.121	19.24555	2
2550	75.77	1	0	20.5	1	1000	0	72.6	1.651	1.61	26.63437	2
2557	75.01	1	0	40	1	1000	0	76	1.69	1.12	26.60971	3
2559	68.99	0	0	44	1	0	0	82	1.6	1.42	32.03125	2
2562	68.23	1	0	28.1	0	500	0	62.7	1.684	1.26	22.10973	3
2569	69.17	0	0	51.25	0	200	0	65.4	1.7	1.44	22.62976	2
2571	87.66	0	0	9.5	0	0	0	58	1.625	1.69	21.9645	0
2573	64.28	1	1	40	1	1000	0	65.7	1.725	2.38	22.07939	2
2575	64.38	0	1	48	1	0	0	61.9	1.618	1.58	23.64469	2
2580	82.79	1	0	185	1	640	0	81.2	1.64	2.42	30.19036	0
2581	80.83	0	0	50	0	800	0	51	1.53	0.84	21.78649	2
2583	72.71	1	0	78	1	0	0	72.45	1.687	1.5	25.45706	2
2592	86.16	1	0	50	1	1000	0	72.2	1.665	2.19	26.04406	1
2596	68.18	0	0	50	0	500	0	38.7	1.58	0.72	15.50232	3
2604	78.98	1	0	54.75	1	1000	0	72.3	1.85	1.63	21.12491	2
2611	71.35	1	0	44	1	0	0	73.6	1.684	2.05	25.95336	2
2613	76.58	1	0	2.7	0	200	0	74.3	1.745	2.37	24.40046	1
2615	51.37	1	1	29.25	1	1000	0	75.9	1.65	2.26	27.87879	2
2617	72.88	1	0	45	1	1000	0	81.4	1.885	0.67	22.90877	4
2621	75.74	1	0	18.75	0	1000	0	70.9	1.67	1.4	25.42221	2
2628	62.73	0	0	38.5	0	0	0	106	1.675	1.58	37.78125	2
2629	56.95	0	1	35.7	1	0	1	90.9	1.702	1.5	31.37941	2
2632	57.49	1	0	82.5	0	800	0	74.9	1.786	2.02	23.48112	2
2633	69.86	1	0	57	1	500	0	83.3	1.74	2.85	27.51354	0
2635	60.32	1	0	43	1	0	0	99.4	1.774	1.27	31.58487	3
2636	85.54	1	0	72	0	1000	0	83.7	1.75	2.33	27.33061	1
2639	68.01	1	0	55	0	1000	0	82.9	1.78	1.93	26.16463	2
2642	70.44	1	1	39.25	1	0	0	76	1.678	1.84	26.99167	2

2643	78.09	1	0	30	0	1000	0	56.5	1.71	0.79	19.32218	3
2645	67.6	1	1	53	1	1000	1	55.15	1.721	1.43	18.62018	3
2648	73.8	0	0	73.5	0	200	0	64.9	1.445	0.9	31.082	2
2650	69.4	1	1	13.75	1	400	0	89.3	1.815	1.38	27.10804	3
2651	73.5	1	0	52	1	800	0	65	1.65	0.93	23.87512	3
2652	65.62	1	0	45		1000	0	77.5	1.8	2.04	23.91975	2
2653	71.29	1	1	13.75	1	0	1	80.5	1.75	1.32	26.28572	3
2655	65.38	1	0	30	1	0	0	70.15	1.7	1.12	24.27336	3
2656	83.85	1	0	70	0	0	0	72.9	1.69	0.63	25.52431	4
2659	63.46	1	1	60	1	1000	0	76.7	1.736	1.24	25.45048	3
2662	70	0	0	4.2	1	1000	0	53.6	1.58	0.92	21.47092	3
2663	73.35	0	0	47	1	400	0	41	1.61	0.62	15.81729	3
2664	67.13	0	0	25	0	1000	0	68.7	1.613	0.94	26.40511	3
2665	65.35	0	0	72	1	1000	0	83.3	1.475	1.05	38.28785	
2668	82.77	1	1	69	1	0	0	74	1.64	0.9	27.51339	3
2672	85.33	0	0	39.75	0	400	0	55.45	1.485	0.75	25.14483	2
2678	72.75	1	1	51.6	1	1000	0	88.1	1.68	1.62	31.21457	2
2680	73.42	1	0	77.5	0	0	0	73	1.744	2.53	24.00103	0
2681	72.97	1	0	58.75	0	1000	0	74.2	1.615	1.23	28.44847	2
2687	77.17	1	0	40	1	1000	0	79	1.68	2.24	27.99036	1
2691	71.65	1	0	46	0	200	0	75.5	1.81	2.43	23.0457	2
2692	71.81	1	1	41.25	1	1000	0	82.2	1.72	1.9	27.78529	2
2697	80.44	0	0	84	0	0	0	65.4	1.67	1.35	23.45011	2
2698	67.14	1	0	72	0	0	0	96.9	1.87	2.36	27.71026	2
2699	74.74	1	0	38	0	1000	0	83.7	1.69	2.2	29.30569	1
2701	65.51	1	1	50.5	0	1000	0	71.8	1.678	2.62	25.50002	1
2705	66.85	1	0	75.25	1	1000	0	102.8	1.88	0.34	29.08556	4
2712	65.69	0	0	35	0	200	0	61.7	1.48	0.78	28.16837	3
2713	50.54	0	0	36	0	0	0	83.9	1.66	1.94	30.44709	2
2714	73.06	1	1	34.2	0	0	0	77.4	1.64	0.96	28.77751	3
2717	82.74	1	0	33	0	500	1	96.3	1.81	2.06	29.39471	2
2718	60.88	1	1	41	1	1000	0	60.1	1.7	1.39	20.79585	3
2721	72.45	0	0	11.25	0	0	0	50.4	1.624	2.09	19.10991	1
2725	72.22	1	0	53	0	1000	0	77	1.65	1.51	28.28283	2

2726	82.98	1	0	27.5		400	0	90.5	1.693	1.44	31.57437	2
2733	69.08	1	1	58	0	0	0	71.6	1.631	2.12	26.91566	1
2737	71.33	1	0	180	0	200	0	108.6	1.7	1.42	37.57785	2
2738	62.52	1	1	48	1	0	0	66	1.78	2.55	20.8307	2
2740	74.5	0	0	51	0	0	1	65.9	1.583	0.63	26.29806	3
2742	70.5	1	0	28.8		500	0	67	1.72	1.18	22.64738	3
2743	78.63	0	1	52.5	1	1000	0	77.35	1.644	1.1	28.61914	2
2744	68.85	1	0	92.75	0	0	0	71.95	1.824	1.68	21.62625	2
2745	74.9	1	0	52	1	400	0	83.7	1.7	1.72	28.96194	2
2746	75.06	0	0	25	0	0	0	78.7	1.728	1.69	26.35647	2
2747	74.77	0	0	64	0	250	0	64.05	1.525	1.07	27.54099	2
2748	80.81	1	1	32	1	0	0	93.4	1.72	1.72	31.57112	2
2749	71.99	1	0	57	0	400	0	96.15	1.73	2.14	32.12603	2
2750	34.21	1	0	102	0	1000	0	89.7	1.65	1.1	32.94766	3
2751	66.83	1	0	9	1	0	0	82.95	1.706	1.7	28.50088	2
2752	70.98	0	1	60	1	400	0	61.5	1.59	0.8	24.32657	3
2753	74.05	1	0	42.4	0	0	0	106.8	1.69	1.98	37.39365	2
2754	69.22	0	1	54	0	0	0	55.6	1.65	2.03	20.42241	1
2755	65.25	1	0	26.4	0	400	0	63.95	1.618	1.68	24.42775	2
2756	62.44	1	1	43	0	1000	0	57.25	1.7	1.02	19.80969	3
2757	80.98	1	0	30	0	800	0	93.9	1.72	1.57	31.74013	2
2758	68.18	0	1	65	1	0	0	50.6	1.562	0.78	20.73903	3
2759	48.11	1	0	10.5	0	1000	0	145.1	1.787	2.66	45.43791	2
2760	76.2	1	0	147.5	1	800	0	87.7	1.793	1.29	27.27966	3
2761	64.59	0	0	47	1	800	0	105.9	1.64	1.48	39.35529	1
2763	56.44	0	0	44	0	1000	0	65	1.68	2.34	23.03005	1
2764	68.53	1	0	55.5	0	1000	0	105.2	1.772	1.3	33.50336	3
2765	70.66	0	0	32	0	800	0	58.9	1.695	1.08	20.50104	3
2766	67.3	1	0	46	1	800	0	83.3	1.624	1.8	31.58443	2
2768	69.21	1	1	44.25	1	0	0	70	1.713	1.7	23.85521	2
2769	71.28	1	0	112	1	800	0	68	1.74	0.9	22.46003	3
2770	68.08	1	0	66.25	1	1000	0	85.6	1.66	1.18	31.06402	3
2771	67.84	1	0	60	1	0	0	58.65	1.64	2.16	21.80622	1
2772	64.26	0	0	28	1	1000	0	88.8	1.612	1.78	34.17299	1

2773	72	1	0	37	0	0	0	110.6	1.818	2.53	33.46319	2
2774	62.94	0	0	39	0	0	0	67.2	1.62	1.85	25.60585	1
2775	67.3	1	0	88	0	1000	0	97.5	1.84	2.41	28.79844	2
2776	82.27	1	1	32	1	400	0	75.75	1.696	1.4	26.33485	2
2777	76.44	0	1	57	1	0	0	41.45	1.576	0.67	16.6883	3
2778	72.55	1	0	75	0	0	0	93.3	1.704	1.95	32.13235	2
2779	61.75	0	1	40	1	400	0	63.2	1.67	1.17	22.66127	3
2780	58.41	0	0	25	0	200	0	66.6	1.74	2.15	21.99762	2
2781	68.73	0	0	10.725	0	0	0	84.4	1.625	0.91	31.96213	3
2782	79.87	1	1	49	1	22	0	76.4	1.76	1.02	24.66426	3
2783	86.45	0	0	17	0	0	0	54.3	1.59	1.29	21.47858	0
3014	72.79	1	0	156	1	1000	0	77.5	1.68	0.66	27.4589	4

Ethics approval was granted by the London-Hampstead Ethics Committees (REC reference 09/H0720/8)

Inclusion criteria:

1. Age >45 years
1. Male or female
2. FEV1/FVC < 0.7
3. Smoking ≥10 pack years
4. No other known established respiratory condition

Exclusion criteria:

1. Other significant respiratory disease
2. Active malignancy
3. Immunosuppression
4. Unable to give informed consent

8.2 Sputum processing

PROTOCOL 1 FOR PROCESSING SPUTUM

Process at 4°C within 2hr of expectoration



Select sputum (if necessary using inverted microscope)



Select out sputum plugs from the saliva, aiming for 0.5-1g



Transfer sputum plugs into an empty (pre-weighed) 15mL polypropylene centrifuge tube



Calculate weight of sputum to be processed (weight of tube and sputum – weight of the empty tube)



Add 8 volumes x sputum weight (in grams) phosphate buffered saline (PBS) and approximately 0.5mL of glass beads



Vortex sputum tube for 15 seconds and then a further 15 minutes on a bench rocker.
Re-vortex for a further 15 seconds



Sub-aliquot 500µl samples into separate Eppendorf tubes. Label with specimen type (Micro and virology), patient code, date of sample and store at -80°C

8.3 Viral RNA extraction

Current to date protocol to extract RNA from sputum at the airways and infectious diseases division Imperial College London, Respiratory Medicine, 3rd floor, Norfolk Place, W2 1PG.

Extraction as per:

Qiamp Viral RNA mini Handbook 03/2018:

Check that Buffer AW1 and Buffer AW2 have been prepared according to the instructions. Buffer AW1 is supplied as a concentrate. Before using for the first time, add the appropriate amount of ethanol (96–100%) as indicated on the bottle

Add carrier RNA reconstituted in Buffer AVE to Buffer AVL according to instructions.

Add 310 µl Buffer AVE to the tube containing 310 µg lyophilized carrier RNA to obtain a solution of 1 µg/µl. Dissolve the carrier RNA thoroughly, divide it into conveniently sized aliquots, and store it at –30 to –15°C. Do not freeze–thaw the aliquots of carrier RNA more than three times. Check Buffer AVL for precipitate, and if necessary incubate at 80°C until the precipitate is dissolved. Calculate the volume of Buffer AVL–carrier RNA mix needed per batch of samples by selecting the number of samples to be simultaneously processed from Table 1 (page 18 of handbook). For larger numbers of samples, volumes can be calculated using the following sample calculation:

$$n \times 0.56 \text{ mL} = y \text{ mL}$$

$$y \text{ mL} \times 10 \text{ µl/mL} = z \text{ µl}$$

Where: **n** = number of samples to be processed simultaneously

y = calculated volume of Buffer AVL

z = volume of carrier RNA–Buffer AVE to add to Buffer AVL

Gently mix by inverting the tube 10 times. To avoid foaming, do not vortex.

Procedure:

1. Pipet 560 µl prepared Buffer AVL containing carrier RNA into a 1.5 ml microcentrifuge tube.

If the sample volume is larger than 140 µl, increase the amount of Buffer AVL–carrier RNA proportionally (e.g., a 280 µl sample will require 1120 µl Buffer AVL–carrier RNA) and use a larger tube.

2. Add 140 µl plasma, serum, urine, cell-culture supernatant or cell-free body fluid to the Buffer AVL–carrier RNA in the microcentrifuge tube. Mix by pulse-vortexing for 15 s. To ensure efficient lysis, it is essential that the sample is mixed thoroughly with Buffer AVL to yield a homogeneous solution. Frozen samples that have only been thawed once can also be used.

3. Incubate at room temperature (15–25°C) for 10 min.

Viral particle lysis is complete after lysis for 10 min at room temperature. Longer incubation times have no effect on the yield or quality of the purified RNA.

4. Briefly centrifuge the tube to remove drops from the inside of the lid.

5. Add 560 µl ethanol (96–100%) to the sample, and mix by pulse-vortexing for 15 s. After mixing, briefly centrifuge the tube to remove drops from inside the lid. Insert a QIAamp Mini column into the VacConnector on the QIAvac 24 Plus vacuum manifold. Use only ethanol, since other alcohols may result in reduced yield and purity of the RNA. Do not use denatured alcohol, which contains other substances such as methanol or methylethylketone. If the sample volume is greater than 140 µl, increase the amount of ethanol proportionally (e.g., a 280 µl sample will require 1120 µl ethanol). To ensure efficient binding, it is essential that the sample is mixed thoroughly with the ethanol to yield a homogeneous solution. The collection tube from the blister pack can be saved for the centrifugation in step 14.

6. Make sure that the main vacuum valve (between the vacuum pump and the vacuum manifold) and the screw cap valve (on the end of the QIAvac 24 Plus vacuum manifold) are closed. Switch on the vacuum pump by pressing the power switch. The vacuum is applied only to the connecting system (if used) and not to the vacuum manifold.

Note: For fast and convenient release of the vacuum pressure, the QIAvac Connecting System or the Vacuum Regulator should be used.

7. Carefully apply 630 µl of the lysate from step 5 into the QIAamp Mini column without wetting the rim. Avoid touching the QIAamp Mini column membrane with the pipette tip.

8. Open the main vacuum valve. Be sure to leave the lid of the QIAamp Mini column open while applying vacuum. After all lysates have been drawn through the QIAamp Mini column, close the main vacuum valve and open the screw cap valve to vent the manifold. Close the screw cap valve after the vacuum is released from the manifold. After closing the main vacuum valve, the vacuum is applied only to the connecting system (if used) and not the vacuum manifold.

Note: If the lysates from individual samples have not completely passed through the membrane despite the VacValves of all other QIAamp Mini columns being closed, place the QIAamp Mini column into a clean 2 mL collection tube (not provided), close the cap, and centrifuge at full speed for 3 min or until it has completely passed through. Continue with step 7 of the spin protocol (page 27). Additional collection tubes can be purchased separately. Centrifugation is performed at 6000 x *g* (8000 rpm) to limit microcentrifuge noise. Centrifugation at full speed will not affect the yield or purity of viral RNA.

9. Repeat steps 7 and 8.

If the sample volume was greater than 140 µl, repeat these steps until all of the lysate has been drawn through the QIAamp Mini column.

10. Apply 750 µl Buffer AW1 to the QIAamp Mini column without wetting the rim.

Avoid touching the QIAamp Mini column membrane with the pipette tip. It is not necessary to increase the volume of Buffer AW1 even if the original sample volume was larger than 140 µl.

11. Open the main vacuum valve. After all Buffer AW1 has been drawn through the QIAamp Mini column, close the main vacuum valve and open the screw cap valve to vent the manifold. Close the screw cap valve after the vacuum is released from the manifold.

12. Apply 750 µl Buffer AW2 to the QIAamp Mini column without wetting the rim. Avoid touching the QIAamp Mini column membrane with the pipette tip. Leave the lid of the column open.

13. Open the main vacuum valve. After all Buffer AW2 has been drawn through the QIAamp Mini column, close the main vacuum valve and open the screw cap valve to vent the manifold. Close the screw cap valve after the vacuum is released from the manifold.

14. Close the lid of the QIAamp Mini column. Remove it from the vacuum manifold and discard the VacConnector. Place the QIAamp Mini spin column in a clean 2 mL collection tube saved from step 5, and centrifuge at full speed for 1 min to dry the membrane completely.

15. Place the QIAamp Mini spin column into a clean 1.5 mL microcentrifuge tube (not provided). Discard the collection tube containing the filtrate. Carefully open the QIAamp Mini spin column. Add 60 µl Buffer AVE equilibrated to room temperature. Close the cap, and incubate at room temperature for 1 min. Centrifuge at 6000 x g (8000 rpm) for 1 min. A single elution with 60 µl Buffer AVE is sufficient to elute at least 90% of the viral RNA from the QIAamp Mini spin column. Performing a double elution using 2 x 40 µl Buffer AVE will increase yield by up to 10%. Elution with volumes of less than 30 µl will lead to reduced yields and will not increase the final concentration of RNA in the eluate. Viral RNA is stable for up to one year when stored at –30 to –15°C or at –90 to –65°C.

Our sputum samples processed essentially the same. Swabs are stored dry in 5mL bijoux at -80 freezer. After taking those out from freezer, add 2-3mL of PBS or RNA/DNA nuclease free water. Leave them on ice to soak well, vortex vigorously time to time. Use 140mkl for RNA extraction using QIAamp Viral RNA Mini Kit according to the manufacturer's instructions as above with the following changes: point 2 – instead of 140 µl plasma we are adding 140 µl swabs; point 10 – instead of adding 60 µl Buffer AVE and Centrifuge at 8000rpm we are adding 23µl Buffer AVE and Centrifuge at 14000rpm.

Reverse transcription and PCR

5µl RNA was converted to cDNA using OL27 as primer and the following RT mix: 2µl of 5xFirst-Strand Buffer (250mM Tris-HCl, 375mM KCl, 15mM MgCl₂), 1µl 0.1M DTT, 0.2µl dNTPs (100mM dATP, 100mM dCTP, 100mM dGTP, 100mM dTTP, 100mM dUTP.) 1.55µl OL27 (25µmolar, Invitrogen) and 0.25µl (10 000U) Reverse Transcriptase (Invitrogen). This cDNA was then used to detect picornaviruses in a single round PCR of 32 cycles¹⁸⁵. If increased sensitivity is required, a heminested second round can be performed using 20-32 further cycles. Differentiation of rhinoviruses from enteroviruses is achieved by restriction enzyme digestion of the PCR product¹⁸⁶.

On the remaining 18µl RNA, RT was then performed as follows: 6.5µl of UHQ (nuclease-free water, Promega) with 2.5µl of random hexamer (0.5µg/µl, Promega) was

added to re suspended RNA and denatured at 70°C for 10 min. After quenching on ice, 6µl UHQ, 10µl of 5x First-Strand Buffer (250mM Tris-HCl, 375mM KCl, 15mM MgCl₂), 5 µl 0.1M DTT, 1.25µl dNTPs, 2µl (200U) of Reverse Transcriptase were added. The mixture was incubated at 37°C for 60 min to yield cDNA. 4 µl of this cDNA was then used for each PCR in a panel of RT-PCR assays based on previously published methods. PCR assays and cycle numbers used were: respiratory syncytial viruses (RSV) A and B nested PCR 30 and 25 cycles¹⁸⁷, coronaviruses OC43 and 229E, nested PCR 30 and 25 cycles¹⁸⁸, influenza viruses A and B nested PCR 30 and 30 cycles¹⁸⁹, parainfluenza viruses 1-3, single round PCR 40 cycles¹⁹⁰, adenoviruses single round PCR 40 cycles¹⁹¹, *Chlamydia pneumoniae* nested PCR 30 and 25 cycles¹⁹² and *Mycoplasma pneumoniae*, single round PCR 40 cycles¹⁹³.

PCR products were fractionated on 2% agarose gels and visualized by ethidium bromide staining and photographed. Any visible band of appropriate size was taken as a positive result. Positive and negative control samples were included in all PCR assays.

TATIANA KEBADZE 2001-2014

RESPIRATORY VIRUS PCR'S

NOTE: FOR USE ON 2720 Thermal Cycler form Applied Biosystems

ALL PRIMERS ARE 25µM STOCK (Invitrogen)

GoTaq Flexi DNA POLYMERASE (PROMEGA)

1. PICORNAVIRUS PCR

10µl OL 27 cDNA
5µl Buffer(10x Buffer)
3µl Mg Cl₂ (1.5□m)
0.85µl dNTP (3.4µm)
3µl OL26 (1.5□m)
3µl OL27 (1.5□m)
0.85µl TAQ(2000U)
24.3µl UHQ

↓

94C 120s
94C 30s)
50C 30s) 32 cycles
72C 120s)

Product size:
Picorna 380bp

72C 240s
4C "forever"

8.4 COPD full data tables

Exacerbation vs Baseline

peak_id	p vlaue	q value	logFC	Retention time	m/z base peak	Compound*
12	0.000416	0.030634	-0.74456	3.92	40	unknown
106	0.001447	0.000001	1.104195	7.28	40	unknown
108	0.00071	0.03925	-1.01599	7.32	56	unknown
128	0.000407	0.031733	-0.68987	7.70	41	2-ethyl-hexanal
150	4.44E-05	0.029451	-0.54597	8.42	67	1,4 Pentadiene
464	2.75E-04	0.033126	-0.70058	18.40	85	Fufuryl alcohol, tetrahydro- 5methyl-cis-
502	0.000867	0.046024	-0.44918	19.62	55	Poor match (MW 226)
597	0.001449	0.000001	-0.63345	22.80	55	Cyclopentanone
625	0.000646	0.038965	-0.83319	24.01	96	Poor match (MW 187)
799	0.0006	0.037885	-0.71174	30.83	111	Poor match (MW 334)
822	0.000384	0.03401	-0.63662	31.83	97	Poor match (MW 138)
843	0.000234	0.034535	-0.39299	32.56	96	Poor match (MW 138)
850	0.000355	0.033622	-0.45166	32.82	97	Poor match (MW 152)
856	0.000302	0.030859	-0.7721	32.97	140	Poor match (MW 288)
858	0.001334	0.000001	-0.58892	33.01	97	Poor match (MW 140)
864	0.001146	0.047525	-0.39011	33.37	83	Ethyl cyclohexane
875	0.000286	0.03161	-0.76974	33.58	71	Poor match (MW 226)
879	0.001032	0.045669	-0.8661	33.64	95	Poor Match (MW 152)
893	0.000914	0.041812	-0.78594	33.94	91	Poor Match (MW 132)
905	0.001157	0.046532	-0.33324	34.13	108	Poor Match (MW 108)
923	0.001325	0.000001	-0.66802	34.51	152	Poor match (MW 152)
942	0.000136	0.036035	-1.11159	35.05	97	Poor match (MW152)
944	0.000214	0.035496	-0.54499	35.09	119	Poor match (MW134)
948	7.12E-05	0.031476	-0.72388	35.20	57	Poor match (MW226)
957	5.41E-04	0.035889	-1.10379	35.43	95	Trans-9-Methyldecalin
960	0.000438	0.030569	-0.68188	35.48	69	Poor match (MW 168)
967	0.00089	0.043765	-0.69419	35.78	69	Poor match (MW 198)
975	0.000907	0.042995	-0.51528	35.99	95	Poor match (MW 566)
979	0.001405	0.000001	-0.5437	36.21	95	Poor match (MW 152)
986	0.000195	0.036903	-0.81992	36.33	97	1-methyl-2-Pentyl-cyclohexane
990	0.000867	0.044257	-0.38592	36.48	85	Poor match (MW 126)
1003	0.000158	0.035051	-0.67616	36.88	97	Poor match (MW 166)
1012	0.001112	0.047611	-0.54022	37.21	166	Poor match (MW 166)
1040	0.000242	0.032098	-0.6745	38.28	149	Diethyl Phthalte
1041	4.12E-05	0.000001	-0.6276	38.37	57	2,9-dimethyl-undecane
1045	3.93E-04	0.032631	-0.53519	38.43	81	Poor match (MW 166)
1298	0.000682	0.039368	-0.63204	46.95	191	2,5-di-tert-butyl phenol
1319	7.48E-05	0.024812	-0.44363	47.44	55	Poor match (MW 300)

RV positive Vs RV negative (irrespective of clinical state)

Peak	p value	q value	Log change	Retention time	M/z base peak	*compound
3	0.005332	0.043677	-1.53813	3.12	207	siloxane
5	0.001603	0.029962	-1.94294	3.49	40	Unknown compound
8	0.000742	0.029844	-1.66787	3.67	40	Unknown compound
54	0.001075	0.027444	-1.88804	5.54	43	1,1,1-trifluoro-2-propanone
59	0.004261	0.040099	-0.85599	5.60	56	Poor match (MW142)
61	0.001018	0.028145	-3.20293	5.64	41	Poor match (MW73)
74	0.002789	0.036646	-1.02526	6.15	77	Trimethylsilyl-fluoride
78	0.000572	0.028123	-1.4237	6.25	83	Poor match (MW264)
123	0.000339	0.024988	-0.83708	7.61	70	Poor match (MW187)
128	0.004387	0.040998	-1.44062	7.70	41	Poor match (MW128)
150	0.00298	0.036281	-0.95402	8.42	67	1,4-Pentadiene
165	0.005129	0.043078	-0.9791	8.72	42	Poor match (MW42)
167	0.003568	0.037583	-0.83604	8.79	56	Poor match (MW56)
186	0.00075	0.029287	-0.72618	9.58	76	Carbon disulfide
194	0.00426	0.040377	-1.7519	9.78	43	cyclobutylamine
250	0.002961	0.037422	-2.96625	11.64	41	methacrolein
261	0.004086	0.040169	-1.26002	12.11	121	Poor match (MW166)
279	0.000951	0.028691	-1.48215	12.72	75	Trimethyl-silanol
304	0.003343	0.037598	-0.9816	13.42	79	Poor match (MW158)
312	0.004393	0.040769	1.860623	13.72	81	Poor match (MW96)
315	0.000152	0.040447	-2.00963	13.85	57	Poor match (MW88)
331	0.001006	0.028406	-1.98761	14.17	29	Poor match (MW178)
333	0.000397	0.022894	-0.58032	14.17	78	Unknown compound
338	0.005072	0.043148	-0.99199	14.55	147	Hexamethyl-disiloxane
343	0.000949	0.029293	-1.38685	14.77	33	Poor match (MW74)
355	0.000853	0.031445	-0.78652	15.12	78	Poor match (MW78)
356	0.006103	0.047363	0.748978	15.24	62	Poor match (MW62)
363	0.00172	0.030428	-0.7638	15.32	56	Poor match (MW98)
370	0.005315	0.044083	-2.06657	15.56	60	Acetic acid
371	0.00298	0.036617	-0.87125	15.61	60	Poor match (MW91)
374	0.001756	0.030265	-1.98394	15.65	84	Poor match (MW84)
380	0.000206	0.030357	-1.93028	15.89	70	2 Butenal
392	0.002975	0.037241	-1.60821	16.12	39	Poor match (MW101)
432	0.002205	0.03216	-0.72504	17.31	44	Pentanal
437	0.002977	0.036921	-1.12047	17.65	70	Poor match (MW112)
438	0.004548	0.040776	-1.85292	17.65	181	Poor match (MW282)
439	0.000694	0.029725	-2.25824	17.70	103	Poor match (MW160)
441	0.000245	0.024995	-2.08729	17.85	131	Poor match (MW160)
451	0.002883	0.036789	-0.69457	18.05	70	Poor match (MW112)
485	0.001697	0.030436	-2.07184	19.19	70	Poor match (MW70)

491	0.005737	0.045589	-1.10262	19.28	111	Poor match (MW249)
498	0.003996	0.03987	-1.36849	19.45	45	Acetoin
499	0.002168	0.031969	-1.28717	19.49	211	Poor match (MW211)
503	0.000286	0.022317	-2.12405	19.72	73	Poor match (MW352)
504	0.001985	0.030991	-2.20151	19.72	74	Poor match (MW134)
511	0.001223	0.027974	-0.81625	19.90	70	Trans-1-Butyl-2methylcyclopropane
513	0.001619	0.029848	-0.84614	20.02	91	Toluene
514	0.002061	0.031429	-0.60618	20.12	85	Octane
519	0.004915	0.04263	-0.66167	20.25	83	Poor match (MW112)
520	0.005394	0.043914	-1.50259	20.25	55	Poor match (MW112)
522	0.001796	0.02979	-1.98905	20.31	59	Poor match (MW120)
524	0.002616	0.036156	-1.56575	20.31	45	Poor match (MW106)
535	0.003398	0.03758	-0.93925	20.78	207	hexamethylcyclotrisiloxane
555	0.000359	0.025095	-1.22887	21.48	153	Poor match (MW252)
558	0.00106	0.027579	-0.84218	21.67	166	tetrachloroethylene
567	0.003236	0.037014	-0.87498	21.89	43	Poor match (MW100)
577	0.001958	0.03168	-0.78027	22.20	56	Poor match (MW100)
581	0.000898	0.030569	-1.90681	22.24	45	1.2-Propanediol
597	0.000159	0.035153	-1.8299	22.80	55	Poor match (MW74)
612	0.002166	0.032295	-0.9313	23.54	51	Branched hexane
613	0.001265	0.027976	-1.52675	23.54	94	Poor match (MW150)
617	0.005317	0.043823	-1.92874	23.70	60	Butanoic acid
625	0.000613	0.029039	-1.81484	24.01	96	Poor match (MW187)
629	0.001564	0.02965	-2.03849	24.07	57	Poor match (MW57)
630	5.57E-05	0.03696	-1.51744	24.13	131	Poor match (MW344)
639	2.76E-04	0.022878	-1.56154	24.34	87	Poor match (MW116)
643	0.006118	0.047201	-1.74484	24.38	56	Poor match (MW99)
649	0.00232	0.033466	-0.65717	24.79	57	Nonane
650	0.001041	0.027628	-0.76442	24.79	71	Poor match (MW200)
656	0.003027	0.036183	-1.17904	25.23	96	Poor match (MW96)
657	0.005192	0.043335	-1.36691	25.53	109	Poor match (MW124)
659	0.004399	0.040536	-0.99744	25.38	82	3 methyl-furan
667	3.53E-05	0.046817	-1.4985	25.73	102	Phenylethyne
683	1.31E-03	0.027633	-1.07968	26.35	72	3-methyl-1acetyl-cyclopentanol
684	0.001433	0.028816	-0.82148	26.35	57	6,6-dimethyl undecane
687	0.002132	0.032147	-1.02512	26.50	58	2-heptanone
694	0.001433	0.028386	-0.8649	26.85	70	Heptanal
695	0.00406	0.040205	-0.91927	26.85	45	Poor match (MW85)
697	0.002806	0.036508	-2.07187	26.95	72	Poor match (MW211)
703	0.001301	0.027847	-1.8246	27.08	69	Poor match (MW238)
721	0.001628	0.029592	-2.46032	27.78	60	Poor match (MW114)
740	0.003807	0.039165	-1.20118	28.42	96	Poor match (MW96)
746	0.001537	0.02956	-0.79076	28.59	91	Propyl-benzene
753	0.001459	0.028463	-0.71689	28.87	105	Poor match (MW120)

789	0.00019	0.031598	-2.55842	30.33	105	1,2,4-trimethyl-benzene
791	0.000659	0.030167	-0.7816	30.54	106	Benzaldehyde
792	0.000969	0.028565	-1.47255	30.63	57	Poor match (MW99)
800	0.000248	0.023474	-0.85565	30.97	56	Poor match (MW128)
805	0.000659	0.029161	-0.82987	31.12	68	D-limonene
816	0.000793	0.030062	-1.56887	31.68	59	Poor match (MW148)
826	0.001958	0.031299	-1.13702	31.94	57	Poor match (MW154)
827	7.37E-05	0.032612	-1.02245	32.03	103	Benzonitrile
828	5.07E-03	0.042874	-0.69162	32.10	57	2-ethyl=1-hexanol
842	0.005572	0.044812	-0.61739	32.50	69	Poor match (MW276)
843	0.000233	0.028149	-1.00738	32.56	96	Poor match (MW138)
844	0.000899	0.029838	-1.41855	32.64	95	Poor match (MW154)
850	0.003398	0.037269	-0.90104	32.82	97	Poor match (MW152)
854	0.003987	0.040084	-1.63057	32.97	81	Poor match (MW200)
856	0.005923	0.046232	-1.37484	32.97	140	Poor match (MW288)
858	0.003236	0.036698	-0.85259	33.01	97	Poor match (MW140)
862	0.000885	0.030895	-1.3819	33.19	94	Phenol
864	0.004469	0.040899	-1.00146	33.37	83	Poor match (MW112)
874	0.000497	0.02539	-1.18278	33.58	56	Poor match (MW184)
875	0.005744	0.045373	-0.95792	33.58	71	Poor match (MW226)
876	0.001179	0.027932	-1.68191	33.58	42	Poor match (MW126)
879	0.001795	0.030152	-1.64341	33.64	95	Poor match (MW152)
885	0.003478	0.037224	-2.27273	33.76	71	Poor match (MW184)
905	0.004991	0.042728	-0.70138	34.13	108	Poor match (MW108)
909	0.001784	0.030344	-1.88754	34.29	43	Poor match (MW172)
912	0.001112	0.027832	-1.68744	34.32	96	Poor match (MW138)
923	0.000234	0.025894	-1.67103	34.51	152	Poor match (MW152)
928	0.002743	0.036402	-0.69729	34.64	57	Nonanal
929	0.000173	0.03273	-1.00415	34.64	105	Poor match (MW182)
936	0.003466	0.037399	-2.02672	34.79	60	Poor match (MW216)
937	0.002653	0.036291	-1.13785	34.93	57	2-methyl-Undecane
940	0.001139	0.027981	-1.00258	34.98	105	Poor match (MW105)
942	0.00086	0.030828	-2.1337	35.05	97	Poor match (MW152)
944	0.001335	0.027673	-1.15852	35.09	119	Poor match (MW134)
945	0.003875	0.039555	-1.62925	35.09	71	Poor match (MW210)
948	0.004191	0.04001	-1.07226	35.20	57	Poor match (MW226)
952	0.003583	0.037435	-1.11364	35.25	163	Poor match (MW193)
960	0.00303	0.035895	-1.3939	35.48	69	Poor match (MW168)
964	0.000225	0.029913	-2.49966	35.66	85	Poor match (MW114)
967	0.000949	0.029973	-1.49664	35.58	69	Poor match (MW198)
970	0.001022	0.027691	-0.91843	35.91	51	Poor match (MW500)
973	0.003963	0.040146	-0.56772	35.91	225	Poor match (MW268)
975	0.003183	0.037379	-0.84653	35.99	95	Poor match (MW566)
977	0.004915	0.042354	-0.61843	36.11	57	dodecane

979	0.000396	0.023867	-1.31374	36.21	95	Poor match (MW152)
982	0.000452	0.025005	-1.22644	36.25	42	Poor match (MW175)
983	0.004093	0.039644	-1.65975	36.25	101	Poor match (MW101)
987	0.005619	0.04492	-1.12157	36.33	97	Poor match (MW166)
990	0.00186	0.030464	-0.81797	36.48	85	Poor match (MW126)
993	0.003686	0.038213	-0.88121	36.56	99	Poor match (MW198)
994	0.000934	0.030238	-0.82729	36.68	130	Poor match (MW398)
995	0.002697	0.036519	-0.96996	36.68	57	Poor match (MW130)
1001	0.005748	0.045134	-0.76218	36.83	71	2,3,5-trimethyl-Decane
1002	0.003564	0.037832	-1.41153	36.83	55	Poor match (MW 252)
1003	0.002482	0.035033	-1.20841	36.88	97	Poor match (MW166)
1004	0.001223	0.0275	-0.90568	36.93	56	Poor match (MW98)
1005	0.001965	0.031047	-1.19776	36.93	73	Poor match (MW 389)
1009	0.000723	0.029977	-1.41702	37.12	55	Poor match (MW 268)
1011	0.003019	0.03642	-0.87159	37.12	108	Poor match (MW 165)
1012	0.001289	0.02805	-1.12914	37.21	166	Poor match (MW 166)
1017	0.002524	0.03525	-0.60685	37.37	149	Poor match (MW 276)
1020	0.001336	0.027271	-1.12664	37.43	105	Poor match (MW327)
1023	0.00447	0.040624	-0.89239	37.61	59	Poor match (MW 59)
1025	0.000989	0.028524	-1.29281	37.65	97	Poor match (MW 97)
1029	0.002883	0.037144	-1.21903	37.71	57	3,5-dimethyl-Undecane
1034	0.003343	0.037282	-0.53958	37.96	57	Decanal
1039	0.003452	0.037552	2.034076	38.15	171	Poor match (MW272)
1041	0.004915	0.042909	-0.95512	38.37	57	Poor match (MW184)
1045	0.002743	0.036769	-0.68023	38.43	81	Poor match (MW166)
1048	0.00118	0.027471	-0.6181	38.55	128	Poor match (MW128)
1051	0.002037	0.031427	-1.2174	38.63	163	Poor match (MW332)
1069	0.004838	0.042802	-0.748	39.22	57	Tridecane
1083	0.004088	0.039891	-2.25769	39.71	218	Poor match (MW424)
1088	0.004838	0.042519	-0.73825	39.82	73	Poor match (MW200)
1100	0.003205	0.037306	-1.60883	40.38	98	Poor match (MW217)
1124	0.00447	0.040347	-0.33443	41.44	135	Poor match (MW236)
1130	0.001159	0.027955	-1.29488	41.64	43	Poor match (MW202)
1141	0.003209	0.037024	-0.70028	41.90	70	Poor match (MW181)
1174	0.005443	0.044042	-0.53299	42.66	69	Poor match (MW186)
1186	0.006206	0.047607	-0.62597	43.07	125	Poor match (MW222)
1228	0.000381	0.02406	-0.82251	44.27	104	Phthalic anhydride
1236	0.001721	0.030043	-0.8108	44.49	91	1-chloro-Decane
1247	0.00234	0.033386	-0.76164	44.88	167	Poor match (MW252)
1250	0.000373	0.02478	-0.20968	44.97	55	1-Dodecanol
1263	0.000488	0.02591	-0.70051	45.31	71	Poor match (MW343)
1269	0.004126	0.039679	-0.5638	45.89	57	Poor match (MW222)
1299	0.004592	0.040896	-1.37079	46.95	57	Poor match (MW212)
1324	0.000275	0.024297	-2.71008	47.76	73	Trimethyl-silane

Exacerbation Virus positive only vs Baseline (negative virus and bacteria)

Peak	p value	FDR	Log fold C	Retention time	m/z base peak	Compound*
16	0.037788	1	-2.83643	4.22	32	Unknown (MW 28)
26	0.030133	1	-1.5118	4.70	39	Unknown (MW 28)
96	0.038892	1	1.235101	6.85	27	Hydrogen cyanide (PM 27)
113	0.011488	1	2.036201	7.37	40	Poor match (MW 32)
136	0.043942	1	0.529689	8.08	45	Poor match (MW 90)
146	0.036927	1	-1.72538	8.34	42	Poor match (MW 174)
148	0.025706	1	-2.08958	8.34	39	Poor match (MW 113)
153	0.049285	1	-2.0421	8.55	71	Poor match (MW 128)
155	0.012466	1	-2.11342	8.61	63	Poor match (MW 28)
166	0.011231	1	-0.59636	8.72	56	Poor match (MW 269)
221	0.0455	1	-0.29889	10.68	43	Poor match (MW 98)
229	0.016252	1	0.757105	11.1	207	Unknown (MW 32)
291	0.026598	1	-1.59032	13.16	61	Poor match (MW 148)
575	0.03402	1	1.445853	22.15	43	Poor match (MW 159)
601	0.0185	1	-1.20925	22.97	43	2,4-dimethyl- Heptane
665	0.018516	1	-0.72101	25.62	54	Poor match (MW 82)
677	0.003288	1	-0.51172	26.15	93	Poor match (MW 93)
680	0.041051	1	-0.53672	26.23	103	1,3,5,7-Cyclooctatetraene
720	0.024965	1	-1.41097	27.71	42	Poor match (MW 72)
725	0.041051	1	-0.55687	27.89	71	Poor match (MW 250)
730	0.012462	1	-1.6499	27.96	112	Poor match (MW 212)
810	0.038322	1	-0.67092	31.29	57	2,6,10-trimethyl-Dodecane
832	0.007058	1	-2.51859	32.16	85	Poor match (MW 282)
876	0.040818	1	-2.08228	33.58	42	Poor match (MW 126)
1060	0.033209	1	-1.99587	38.86	67	Poor match (MW 110)
1133	0.041024	1	-1.15665	41.64	111	Poor match (MW 249)
1165	0.025991	1	-1.62532	42.49	141	Poor match (MW 196)
1201	0.009528	1	2.427933	43.49	149	Poor match (MW 164)
1261	0.000744	1	-3.03076	45.23	85	Poor match (MW 282)

RV positive exacerbation vs RV negative exacerbation

Peak number	p value	Q value	Log change	Retention time	M/z base peak	Compound*
56	0.044572	1	-0.10850539	5.54	60	Methyl formate
84	0.003619	1	-1.51457938	6.38	128	1-en-2-ylether fluoromethyl pentafluoroprop
96	0.043438	1	1.578246034	6.85	27	Poor match (MW 27)
113	0.002237	1	2.1852026	7.37	40	Poor match (MW 40)
145	0.015791	1	-2.72255791	8.34	27	Poor match (MW 27)
155	0.018786	1	-1.78544783	8.61	63	Unknown
166	0.00567	1	-0.92684408	8.72	56	Poor match (MW 269)
225	0.04538	1	1.286099783	10.79	41	Poor match (MW 86)
247	0.035997	1	1.138085204	11.45	57	Poor match (MW 194)
290	0.031851	1	2.634632951	13.16	45	Poor match (MW 88)
293	0.039382	1	-1.88406917	13.22	73	Poor match (MW 116)
344	0.026108	1	2.411487777	14.77	43	2 methyl 1 propanol
385	0.019771	1	-2.37507975	15.95	69	Trifluoro-anhydride Acetic acid
403	0.023635	1	1.536451763	16.31	45	Poor match (MW 76)
466	0.023991	1	0.730945451	18.45	57	Poor match (MW 98)
503	0.034629	1	2.338637236	19.72	73	Poor match (MW 352)
516	0.026046	1	-0.82777771	20.17	97	1-(2-furanyl)- 3-Butene-1. Diol
561	0.000597	1	-3.29118952	21.71	83	Poor match (MW 128)
566	0.039348	1	-0.87397436	21.79	97	1-isocyano-Napthalene
583	0.018788	1	-0.90632326	22.34	97	Poor match (MW 112)
593	0.005029	1	-1.12867986	22.61	111	Poor match (MW 126)
594	0.00183	1	-2.43183806	22.68	55	Poor match (MW 454)
601	0.002044	1	-2.55611946	22.97	43	2,4-dimethyl-Heptane
607	0.032138	1	-0.78505627	23.30	111	1-(2-thienyl)-1-Pentanone
633	0.049372	1	-0.81431821	24.13	57	Acetic acid, 1- [(1dimethylethyl)thio]
643	0.028953	1	1.898477265	24.38	56	Poor match (MW 99)
644	0.014158	1	-0.91579973	24.52	111	Poor match (MW 126)
648	0.014145	1	-0.9899415	24.69	55	Poor match (MW 98)
669	0.045741	1	-0.73634008	25.91	97	Poor match (MW 252)
670	0.014157	1	-0.76393634	25.91	55	Poor match (MW 128)
675	0.022451	1	-1.88772249	26.06	41	Poor match (MW 128)
679	0.008715	1	0.675873729	26.15	169	Poor match (MW 228)
698	0.047925	1	-1.41365346	26.95	97	Poor match (MW 112)
705	0.007978	1	-1.72024285	27.08	71	Poor match (MW 86)
719	0.006481	1	-1.2746258	27.71	55	Cyclohexanone
720	0.00414	1	-1.13971867	27.71	42	Poor match (MW 72)
730	0.006587	1	-1.31552699	27.96	112	Poor match (MW 212)
757	0.004135	1	-1.52605122	29.15	176	Poor match (MW 389)

786	0.045681	1	-1.91975505	30.25	77	Poor match (MW 106)
830	0.045578	1	-1.44917718	32.16	56	Poor match (MW 157)
832	0.000299	1	-3.40546299	32.16	85	Poor match (MW 85)
870	0.015458	1	2.358744368	33.43	97	Poor match (MW 138)
906	0.015844	1	0.557125714	34.13	107	Phosphorous pentafluoride
959	0.045211	1	0.905177123	35.43	177	Poor match (MW 346)
966	0.035972	1	-1.53265275	35.66	43	Poor match (MW 142)
972	0.046112	1	0.204942151	35.91	149	Ethaneperoxoic acid
991	0.045741	1	-0.87666932	36.48	147	Poor match (MW 406)
996	0.010437	1	-2.28830577	36.76	57	Poor match (MW 139)
998	0.005257	1	-1.8892715	36.76	60	Poor match (MW 211)
1000	0.022451	1	0.796198455	36.76	44	2-(methylamino)ethyl-Phenol
1039	0.045993	1	-2.03974195	38.15	171	Poor match (MW 272)
1059	0.024225	1	2.284709959	38.86	82	Poor match (MW 182)
1115	0.045741	1	-1.55280367	40.93	197	Poor match (MW 239)
1150	0.030502	1	1.236292846	42.21	57	Poor match (MW 184)
1165	0.030542	1	-1.61787653	42.49	141	Poor match (MW 196)
1190	0.041429	1	-1.74606236	43.17	71	Poor match (MW 216)
1211	0.045489	1	-1.7090798	43.85	59	Poor match (MW 142)
1217	0.048002	1	-1.293735	43.99	151	Poor match (MW 151)
1218	0.032578	1	-1.35535604	43.99	109	Poor match (MW 186)
1230	0.009373	1	1.819896695	44.31	71	Poor match (MW 170)
1247	0.043542	1	0.715629056	44.88	167	Poor match (MW 252)
1257	0.018788	1	-1.58373311	45.18	71	Poor match (MW 170)
1261	0.001072	1	-2.6969458	45.23	85	Poor match (MW 282)
1281	0.033829	1	-2.53461952	46.41	149	Hept-4yl isobutyl ester phthalic acid

Bacterial positive exacerbation vs Baseline Bacterial negative (virus negative)

Peak number	p value	Q value	Log change	Retention time	M/z base peak	Compound
22	0.018575	1	-1.444157859	4.55	69	Poor match (MW 331)
31	0.004226	1	-0.324384953	4.92	58	Acetone
37	0.017468	1	-0.723932552	5.01	29	Poor match (MW 174)
40	0.040785	1	-0.938602722	5.22	41	Poor match (MW 170)
42	0.003911	1	-1.966161017	5.22	27	Tetra-t-butylethane
47	0.012659	1	-0.872286124	5.32	43	Cyclobutylamine
50	0.002729	1	-1.444529972	5.49	20	Poor match (MW 124)
58	0.039237	1	-0.647420428	5.60	41	1-butene
64	0.004444	1	-0.943975309	5.76	43	Dimethyl-Diazene
67	0.016516	1	-1.062846708	5.91	51	Pentafluoroethane
73	0.030317	1	-0.955843764	6.15	44	Hydroxyurea
85	0.047367	1	-1.116356833	6.38	48	Poor match (MW 80)

97	0.005405	1	-2.441232944	6.95	64	Poor match (MW 64)
114	0.044933	1	-0.905467646	7.37	101	Trichloromonofluoromethane
123	0.0475	1	-0.894911952	7.61	70	Poor match (MW 187)
136	0.044036	1	-0.750923204	8.08	45	Poor match (MW 90)
150	0.042384	1	-0.574950977	8.42	67	1,4-Pentadiene
169	0.045742	1	-1.443994588	8.85	58	Propanal
173	0.013844	1	-2.767059891	8.96	75	Unkown
195	0.01962	1	-1.252726581	9.78	41	Poor match (MW 56)
196	0.029685	1	-1.273682851	9.78	39	Poor match (MW 70)
208	0.003721	1	-2.512424783	10.16	84	2-butyl-1methyl-pyrrolidine
211	0.002503	1	-1.763714044	10.23	28	Hydrogen cyanide
215	0.006571	1	-1.748509548	10.39	57	Unkown
219	0.01867	1	-0.891214837	10.68	57	Butyl isobutyl ester sulphurous acid
222	0.009968	1	-0.966763014	10.68	73	Poor match (MW 74)
234	0.037739	1	-0.848099989	11.21	56	n-hexane
244	0.012531	1	-1.71481161	11.45	84	Poor match (MW 220)
258	0.012933	1	-1.328909235	12.06	84	Poor match (MW 84)
259	0.000282	1	-2.306748246	12.06	55	Poor match (MW 114)
274	0.002791	1	-1.9891306	12.43	194	Poor match (MW 237)
290	0.045873	1	-1.967327811	13.16	45	Poor match (MW 88)
317	0.038451	1	-1.004589772	13.89	123	Poor match (MW 220)
340	0.022176	1	-2.040696537	14.64	231	Poor match (MW 310)
351	0.027407	1	0.184987968	15.06	61	Poor match (MW 132)
362	0.018767	1	-0.80383105	15.27	181	Diglycolic acid, di(pentafluorobenzyl) ester
370	0.008761	1	-1.847959715	15.56	60	Acetic acid
375	0.00066	1	-1.175925134	15.65	43	Acetic anhydride
377	0.01104	1	1.912170071	15.70	43	Poor match (MW 132)
387	0.003739	1	1.745927532	15.95	43	Propanedioic acid
391	0.01742	1	-0.773663283	16.12	43	Diacetyl sulphide
399	0.02629	1	-1.362638623	16.25	31	Poor match (MW 100)
424	0.034922	1	-0.80621652	16.83	99	Poor match (MW 196)
443	0.000284	1	-1.877317596	17.85	99	Poor match (MW 168)
457	0.005492	1	-0.27104931	18.25	69	Trifluormethansulfinyl fluoride
458	0.001409	1	-0.726690866	18.25	86	Poor match (MW 220)
475	0.01408	1	-1.215345185	18.76	70	Poor match (MW 168)
482	0.012289	1	-1.424600719	19.11	57	Poor match (MW 116)
483	0.014574	1	-0.777525876	19.11	97	Poor match (MW 154)
484	0.030701	1	-0.120828588	19.11	127	Isopropyl ester Glycine N-trifluoroacetyl
488	0.010266	1	-1.421997481	19.19	32	Poor match (MW 154)
502	0.009146	1	-1.121742575	19.62	55	Poor match (MW 226)
508	0.020473	1	-1.329064146	19.86	98	Toulene-D8

571	0.045672	1	-1.016196283	21.96	57	Poor match (MW 194)
587	0.03623	1	0.86934908	22.43	57	Poor match (MW 126)
611	0.000205	1	-0.532428311	23.49	71	Poor match (MW 256)
622	0.040782	1	1.437884741	23.94	73	Poor match (MW 233)
625	0.00939	1	-1.495014023	24.01	96	Poor match (MW 187)
626	0.024735	1	-1.550138812	24.01	95	Poor match (MW 350)
647	0.019626	1	-0.39995001	24.69	43	Poor match (MW 154)
648	0.016514	1	-1.030149731	24.69	55	Ethenyl ester 2-propenoic acid
672	0.010554	1	-1.285546651	26.01	60	Poor match (MW 60)
675	0.029969	1	-1.567361026	26.06	41	Poor match (MW 128)
705	0.014176	1	-1.827095257	27.08	71	Poor match (MW 86)
713	0.049285	1	1.53983982	27.49	70	Poor match (MW 157)
747	0.036037	1	-0.791641621	28.59	149	Diethyl Phthalate
752	0.023507	1	-1.043993327	28.73	93	Poor match (MW 202)
769	0.016014	1	-1.381835452	29.45	191	Poor match (MW 320)
840	0.009156	1	-0.836367509	32.42	175	Poor match (MW 298)
893	0.014213	1	-1.435083898	33.94	91	Poor match (MW 132)
916	0.011879	1	-1.034805964	34.32	105	Poor match (MW 244)
931	0.034878	1	-1.279561572	34.73	141	Poor match (MW 196)
939	0.032939	1	-1.339908217	34.93	205	1,2,4,6-Trtert-butyl-4methyl-2,5-cyclohexadien-1-one
944	0.008744	1	-1.488559163	35.09	119	Poor match (MW 134)
951	0.040782	1	-0.950630892	35.25	108	Poor match (MW 227)
959	0.020845	1	-0.861421425	35.43	177	Poor match (MW 346)
983	0.024353	1	-1.549212058	36.25	101	Poor match (MW 101)
999	0.0481	1	-0.749397438	36.76	71	Poor match (MW 380)
1028	0.03575	1	-1.543488898	37.65	92	Poor match (MW 148)
1049	0.031593	1	-1.72208245	38.55	129	Poor match (MW 129)
1072	0.017131	1	-1.873316131	39.22	102	Poor match (MW 174)
1096	0.025451	1	-2.577751206	40.30	83	Poor match (MW 222)
1108	0.00218	1	-2.171483336	40.70	58	Poor match (MW 246)
1109	0.042036	1	-1.337677219	40.70	225	Poor match (MW 408)
1136	0.033535	1	-1.287703894	41.69	243	Poor match (MW 354)
1147	0.022757	1	-1.420027382	42.12	135	Poor match (MW 420)
1181	0.042384	1	-1.106394964	42.92	70	Endo-3-acetamidocamphor
1187	0.007348	1	-2.278615281	43.10	221	Poor match (MW 294)
1189	0.004281	1	-1.963668779	43.10	105	Poor match (MW 225)
1200	0.045707	1	1.38044991	43.49	170	N-buetyl Benzenesulfonamide
1206	0.042384	1	-0.389573872	43.70	73	Poor match (MW 284)
1214	0.047504	1	-0.928673558	43.93	154	Poor match (MW 154)
1216	0.023196	1	0.942421832	43.93	138	1-(1-cyclopropyl-pentyl)piperidine
1265	0.03489	1	-0.640573199	45.31	113	Poor match (MW 212)

1271	0.036291	1	-0.654495434	45.89	244	Poor match (MW 244)
1281	0.039237	1	-1.634802668	46.41	149	Hept-4yl isobutyl ester phthalic acid
1319	0.029828	1	-0.641001371	47.44	55	Poor match (MW 300)

RSV positive vs RSV Negative

Peak	p value	q value	log correction	Retention time	M/z base peak	Compound*
76	0.004722	0.895101	2.219529	6.25	131	Poor match (MW28)
96	0.026118	0.962734	-0.87541	6.85	27	Poor match (MW27)
103	0.001639	1	2.876863	7.14	41	Poor match (MW41)
111	0.010671	0.943996	0.276735	7.32	43	Poor match (MW43)
140	0.004137	0.914897	-1.60227	8.27	38	Poor match (MW91)
162	0.01401	0.929596	-0.73627	8.66	39	Poor match (MW488)
180	0.018607	0.949657	-1.19467	9.17	53	Poor match (MW84)
245	0.01206	0.889075	-1.28098	11.45	31	Poor match (MW60)
287	0.009195	0.938629	-2.12971	12.92	75	Tert-butyldimethylsilanol
316	0.016225	0.936101	-1.11717	13.85	71	Poor match (MW170)
319	0.005791	0.768461	-1.64348	13.89	75	Poor match (MW150)
322	0.002875	0.762909	-0.84786	13.94	42	Poor match (MW116)
357	0.002511	0.833179	-1.8475	15.24	39	Poor match (MW39)
407	0.012001	0.936809	-0.37305	16.42	45	Poor match (MW118)
532	0.005667	0.835508	-1.44478	20.65	77	Dimethyl-silandiol
576	0.000502	0.666816	2.74355	22.15	58	Poor match (MW158)
591	0.01285	0.897467	-1.28705	22.57	45	Poor match (MW130)
601	0.026038	0.987199	1.33702	22.97	43	2,4-dimethyl-Heptane
622	0.016559	0.915583	-1.59116	23.94	73	Poor match (MW233)
727	0.002063	0.912655	-1.48144	27.96	56	Poor match (MW99)
747	0.005481	0.909096	0.801017	28.59	149	Diethyl phthalate
750	0.018075	0.95942	-1.10574	28.73	96	Poor match (MW187)
758	0.026029	1	-1.94101	29.23	209	Poor match (MW224)
766	0.02092	0.895504	-1.67729	29.39	209	Poor match (MW 344)
848	0.00728	0.805092	-0.32881	32.64	244	Poor match (MW244)
868	0.011532	0.956465	-0.38067	33.37	149	Poor match (MW402)
921	0.015585	0.94004	1.5493	34.40	83	Poor match (MW182)
973	0.006469	0.780396	-0.25224	35.91	225	Poor match (MW268)
980	0.023866	0.959721	-2.00975	36.21	108	Poor match (MW165)
993	0.019334	0.88471	-0.82246	36.56	99	Poor match (MW198)
1028	0.022091	0.91608	1.620676	37.65	92	Poor match (MW148)
1043	0.010624	1	1.225375	38.87	57	Poor match (MW282)
1058	0.02	0.884652	0.930217	38.86	97	Poor match (MW228)
1127	0.019312	0.91523	-1.27071	41.44	107	Poor match (MW126)
1181	0.018698	0.918957	1.132425	42.92	70	Poor match (MW181)
1198	0.014489	0.915581	-1.37468	43.49	69	Poor match (MW212)

Bacterial colonisation baseline colonised vs Baseline URTF

Peak number	p value	Q value	Log correction	Retention time	M/z peak base	Compound
20	0.028391	1	1.734642	4.39	44	Monoammonium salt carbamic acid
63	0.014089	1	-0.94413	5.71	43	Acetone
116	0.049563	1	2.303382	7.42	104	Unknown
184	0.049813	1	0.622514	9.31	45	Isopropyl Alcohol
234	0.016309	1	1.111628	11.21	56	n-Hexane
235	0.018182	1	1.043471	11.21	57	Sulfurous acid, butyl isobutyl ester
246	0.040037	1	1.746174	11.45	55	Poor Match (MW 114)
250	0.020678	1	4.890652	11.64	41	Methacrolein
256	0.019357	1	1.572062	11.77	55	Unknown
288	0.045328	1	0.637264	13.08	43	2-Butanone
317	0.041144	1	2.18502	13.89	123	Poor Match (MW 138)
323	0.025611	1	2.195828	13.94	80	Poor Match (MW 80)
366	0.027674	1	2.433587	15.40	58	Poor Match (MW 58)
384	0.041186	1	1.221237	15.95	55	Poor Match (MW 112)
387	0.009502	1	-0.73652	15.95	43	Bis(1-methylethyl)ester Propanedioic acid
394	0.044312	1	2.175828	16.22	225	Poor Match (MW 164)
420	0.024642	1	1.009587	16.78	57	Poor Match (MW 72)
428	0.039134	1	1.673657	17.05	43	3,3-Dimethyl-2,4-pentane dione
433	0.006485	1	0.902781	17.38	69	Methyl methacrylate
435	0.013065	1	1.630544	17.49	43	n-Propyl acetate
439	0.0498	1	1.863573	17.70	103	Poor Match (MW 160)
454	0.049563	1	1.741336	18.15	31	2-Chloroethanol
483	0.035094	1	-0.63411	19.11	97	Poor Match (MW 154)
486	0.006196	1	2.746658	19.19	56	Poor Match (MW 112)
488	0.014237	1	1.932242	19.19	55	Poor Match (MW 238)
490	0.033863	1	0.673967	19.28	51	Poor Match (MW 320)
498	0.030639	1	1.541502	19.45	45	Acetoin
503	0.029119	1	1.909535	19.72	73	Poor Match (MW 352)
505	0.029146	1	2.432185	19.82	57	Poor Match (MW 128)
509	0.035571	1	1.698995	19.86	43	Poor Match (MW 98)
511	0.010406	1	0.914283	19.90	70	Trans-1-Butyl-2methylcyclopropane
520	0.037331	1	2.061635	20.25	55	Poor Match (MW 55)
524	0.044669	1	1.613445	20.31	45	Poor Match (MW 106)
530	0.030639	1	0.374189	20.55	55	Poor Match (MW 112)
534	0.045328	1	0.302125	20.72	97	Poor Match (MW 154)

535	0.020241	1	0.839184	20.78	207	siloxane
557	0.022426	1	2.173102	21.52	43	Poor Match (MW 128)
563	0.010056	1	2.27252	21.79	109	Poor Match (MW 124)
564	0.017811	1	2.072438	21.79	55	Poor Match (MW 134)
567	0.049813	1	0.645631	21.89	43	2-Hexanone
570	0.011413	1	2.296676	21.96	43	Poor Match (MW 170)
575	0.009502	1	1.163847	22.15	43	Poor Match (MW 159)
577	0.003948	1	0.744114	22.20	56	Hexanal
584	0.039833	1	1.28046	22.34	97	unknown
597	0.027674	1	0.725837	22.80	55	cyclopentanone
611	0.024943	1	0.350321	23.49	71	Poor Match (MW 256)
618	0.020234	1	1.116446	23.81	75	Poor Match (MW 176)
621	0.009262	1	1.587249	23.87	105	Unknown
626	0.02583	1	2.811596	24.01	95	Poor Match (MW 350)
635	0.040037	1	2.163263	24.19	69	Poor Match (MW 98)
645	0.027512	1	2.587205	24.52	43	Poor Match (MW 234)
659	0.018182	1	1.221605	25.38	82	3-methyl-Furan
671	0.024967	1	0.253275	25.91	70	Poor Match (MW 126)
687	0.013065	1	1.511403	26.50	58	2-Heptanone
691	0.047076	1	1.195905	26.68	94	Poor Match (MW 94)
694	0.027682	1	0.530922	26.85	70	Heptanal
695	0.005731	1	2.22756	26.85	45	Poor Match (MW 85)
696	0.026579	1	2.670689	26.95	43	Poor Match (MW 100)
703	0.006006	1	3.178507	27.08	69	Poor Match (MW 238)
709	0.033863	1	1.015326	27.28	105	Poor Match (MW 120)
711	0.020241	1	1.91902	27.39	87	Poor Match (MW 158)
719	0.0225	1	-1.27212	27.71	55	Poor Match (MW 98)
720	0.041188	1	-0.715	27.71	42	Poor Match (MW 72)
721	0.031737	1	2.649602	27.78	60	Poor Match (MW 114)
731	0.003948	1	1.079801	28.02	249	Poor Match (MW 442)
732	0.012474	1	2.46174	28.02	281	siloxane
740	0.003053	1	1.322582	28.42	96	Poor Match (MW 96)
761	0.045328	1	-0.69771	29.29	93	B-Phellandrene
776	0.010359	1	-1.21764	29.70	43	2-methyl-propanol
785	0.033863	1	-4.084	30.25	105	Poor Match (MW 215)
788	0.049813	1	-0.78207	30.33	93	Poor Match (MW 136)
800	0.010406	1	0.901599	30.97	56	Octanal
805	0.037373	1	0.482622	31.12	68	D-Limonene
822	0.005075	1	1.981889	31.83	97	Poor Match (MW 138)
824	0.048764	1	-1.85148	31.83	45	Poor Match (MW 90)
843	0.020241	1	0.910473	32.56	96	Poor Match (MW 138)
844	0.005065	1	2.377529	32.64	95	Poor Match (MW 154)
850	0.008239	1	0.913566	32.82	97	Poor Match (MW 152)
856	0.032351	1	1.934828	32.97	140	Poor Match (MW 288)

858	0.013065	1	1.088829	33.01	97	Poor Match (MW 140)
862	0.024975	1	1.770217	33.19	94	Phenol
864	0.030639	1	0.685182	33.37	83	Poor Match (MW 112)
869	0.027682	1	1.292312	33.43	109	Poor Match (MW 109)
872	0.03013	1	2.511271	33.51	67	Poor Match (MW 111)
875	0.018157	1	1.454282	33.58	71	Poor Match (MW 226)
876	0.010322	1	3.544205	33.58	42	Poor Match (MW 126)
881	0.015105	1	-2.14013	33.64	57	Poor Match (MW 168)
882	0.017067	1	-2.17614	33.64	52	Poor Match (MW 52)
883	0.0225	1	1.010089	33.71	57	Poor Match (MW 244)
890	0.013065	1	0.873774	33.85	73	Poor Match (MW 74)
893	0.048316	1	2.611844	33.94	91	Poor Match (MW 132)
896	0.020241	1	1.002236	34.01	126	Poor Match (MW 296)
903	0.045328	1	1.241246	34.13	51	Poor Match (MW 180)
905	0.018157	1	0.967769	34.13	108	Poor Match (MW 108)
910	0.020241	1	0.978048	34.29	58	Poor Match (MW 142)
912	0.01306	1	1.843702	34.32	96	Poor Match (MW 138)
915	0.023499	1	2.091234	34.32	43	Poor Match (MW 194)
923	0.030639	1	1.074756	34.51	152	Poor Match (MW 152)
926	0.024954	1	2.185919	34.58	81	Poor Match (MW 182)
928	0.037373	1	0.654327	34.64	57	Nonanal
930	0.010322	1	2.469748	34.73	110	Poor Match (MW 208)
935	0.037373	1	0.543546	34.79	170	4-methyl-Undecane
940	0.033863	1	0.84183	34.98	105	Poor Match (MW 314)
941	0.032351	1	1.328572	34.98	165	Poor Match (MW 165)
942	0.005027	1	3.252459	35.05	97	Poor Match (MW 152)
945	0.040966	1	1.243176	35.09	71	Poor Match (MW 210)
948	0.016309	1	0.706	35.20	57	Poor Match (MW 226)
957	0.039862	1	2.494761	35.43	95	Poor Match (MW 152)
958	0.008239	1	1.148733	35.43	152	Poor Match (MW 152)
960	0.003474	1	2.103411	35.48	69	Poor Match (MW 168)
967	0.000892	1	3.807325	35.78	69	Poor Match (MW 198)
969	0.027682	1	1.386715	35.83	105	silane
970	0.049813	1	0.430973	35.91	51	Poor Match (MW 500)
975	0.013065	1	0.991666	35.99	566	Poor Match (MW 566)
979	0.002679	1	1.755384	36.21	979	Poor Match (MW 152)
983	0.033766	1	2.272378	36.25	101	Poor Match (MW 101)
986	0.023442	1	1.630804	36.33	168	Poor Match (MW 168)
987	0.006075	1	1.783728	36.33	166	Poor Match (MW 166)
989	0.011509	1	1.334414	36.44	207	Poor Match (MW 221)
990	0.049813	1	0.72831	36.48	85	Poor Match (MW 126)
994	0.041188	1	0.678972	36.68	130	Poor Match (MW 398)
997	0.049113	1	1.451956	36.76	83	Poor Match (MW 83)
1002	0.014586	1	1.870891	36.83	55	Poor Match (MW 55)

1003	0.007315	1	1.793309	36.88	97	Poor Match (MW 166)
1005	0.015432	1	2.001585	36.93	73	Poor Match (MW 389)
1009	0.018176	1	2.983841	37.12	55	Poor Match (MW 268)
1010	0.018176	1	1.257907	37.12	91	Poor Match (MW 91)
1012	0.010406	1	1.307021	37.21	166	Poor Match (MW 166)
1017	0.016309	1	0.879882	37.37	149	Poor Match (MW 276)
1021	0.005688	1	1.86195	37.43	163	Poor Match (MW 298)
1023	0.027682	1	0.939597	37.61	59	Poor Match (MW 57)
1027	0.00319	1	-2.50681	37.65	71	Poor Match (MW 114)
1029	0.027682	1	0.991581	37.71	57	3,5-dimethyl Undecane
1037	0.037373	1	1.242501	38.15	71	Poor Match (MW 226)
1040	0.005741	1	2.998916	38.28	149	Diethyl Phthalate
1041	0.030639	1	1.980135	38.73	57	Poor Match (MW 184)
1043	0.027512	1	2.189234	38.73	57	Poor Match (MW 282)
1045	0.013065	1	1.230949	38.43	81	Poor Match (MW 166)
1051	0.02629	1	1.661814	38.63	163	Poor Match (MW 332)
1058	0.033863	1	1.406084	38.86	97	Poor Match (MW 228)
1063	0.049813	1	0.860255	39.01	85	Poor Match (MW 296)
1066	0.037373	1	0.735959	39.08	59	1-(2-butyoxy-1methylethoxy) - Propanol
1069	0.037373	1	0.611507	39.22	57	Tridecane
1075	0.013065	1	1.377101	39.46	71	3-methyl-5-propyl-Nonane
1082	0.03063	1	1.780658	39.71	137	Poor Match (MW 166)
1108	0.043918	1	2.149362	40.70	58	Poor Match (MW 246)
1111	0.035571	1	1.459554	40.83	121	Poor Match (MW 205)
1122	0.009266	1	-2.49256	41.36	91	Poor Match (MW 285)
1123	0.008239	1	0.599003	41.36	184	Poor Match (MW 184)
1124	0.020241	1	1.842601	41.44	135	Poor Match (MW 236)
1163	0.022352	1	1.63509	42.49	73	Poor Match (MW 327)
1164	0.023499	1	-2.06483	42.49	56	Poor Match (MW 187)
1167	0.011668	1	1.992507	42.55	113	Poor Match (MW 113)
1174	0.020241	1	0.997493	42.66	69	Poor Match (MW 186)
1207	0.018182	1	0.944564	43.70	147	Poor Match (MW 366)
1215	0.049375	1	1.304524	43.93	262	Poor Match (MW 262)
1224	0.047513	1	1.445209	44.07	94	Poor Match (MW 150)
1277	0.030603	1	1.430665	46.24	121	Poor Match (MW 380)
1279	0.048784	1	-2.11523	46.24	120	Poor Match (MW 222)
1281	0.041188	1	2.447917	46.41	149	Hept-4-yl isobutyl ester
1285	0.041144	1	0.158618	46.59	57	Poor Match (MW 198)
1288	0.000762	1	-3.09103	46.67	70	Poor Match (MW 127)
1308	0.008223	1	0.800294	47.24	262	Poor Match (MW 361)
1326	0.02411	1	-0.39104	47.96	57	Poor Match (MW 448)

Non Smoker vs Smoker

Peak number	p value	Q value	Log correction	Retention time	m/z base peak	Compound*
58	0.000544	0.036113	-0.619760029	5.60	41	1-Butene
92	0.000103	0.011433	-0.307308335	6.77	91	Toluene
124	0.000169	0.016029	-0.732253108	7.61	55	Poor match (MW 198)
129	0.000508	0.035503	-0.67942873	7.82	42	Methyl-Cyclobutane
130	0.000549	0.034666	-0.804500876	7.82	41	Poor match (MW 170)
150	0.000694	0.036839	-0.73065882	8.42	67	1,4 Pentadiene
157	0.00021	0.018545	-0.525644351	8.61	85	Poor match (MW 159)
189	0.00059	0.034048	-0.289012377	9.62	43	N,N,O-Triacetylhydroxylamine
190	0	0	-1.584766018	9.65	66	Poor match (MW 184)
197	2.15E-07	4.08E-05	-1.293567951	9.85	41	Poor match (MW 170)
198	1.06E-08	2.34E-06	-1.359895019	9.85	40	Poor match (MW 40)
199	8.34E-04	4.10E-02	-0.629314616	9.91	84	Methylene Chloride
262	4.88E-09	1.29E-06	-1.19011921	12.11	67	Poor match (MW 142)
264	3.64E-07	6.04E-05	-0.775381493	12.11	84	Poor match (MW 181)
269	5.11E-07	7.54E-05	-0.774790924	12.28	82	2-methyl-Furan
284	2.22E-04	1.84E-02	-0.585298742	12.85	82	Poor match (MW 280)
295	0.000166	0.016953	-0.712755421	13.30	54	Poor match (MW 55)
298	0	0	-1.612053863	13.35	67	Poor match (MW 172)
320	0	0	-1.946557765	13.89	79	Poor match (MW 158)
338	0.000435	0.033933	-0.653193295	14.55	147	Hexamethyl-disiloxane
408	9.36E-07	0.000124	-0.511339453	16.42	43	N,N,O-Triacetylhydroxylamine
414	4.44E-16	1.47E-13	-1.418501153	16.63	96	2,5-Dimethyl furan

437	4.35E-04	3.21E-02	- 0.629784482	17.65	70	Poor match (MW 112)
502	0.000704	0.03595	- 0.697866531	19.62	55	Poor match (MW 226)
522	0.000591	0.032684	-0.67955119	20.31	58	Poor match (MW 120)
601	6.04E-05	0.007283	- 0.947134272	22.97	43	2,4-dimethyl-Heptane
1041	5.81E-04	0.03504	- 0.636982386	38.37	57	Poor match (MW 184)
1062	0.000835	0.039573	- 0.472186734	38.92	175	1.3(1-1dimethyl)-Benzene

Fatty acid degradation



[illegible]

9 References

1. J.Gibson. Acute lower respiratory tract infections. *European Lung White Book* 2014: chapter 18.
2. Oliver BGG, Lim S, Wark P, et al. Rhinovirus exposure impairs immune responses to bacterial products in human alveolar macrophages. *Thorax* 2008; **63**(6): 519-25.
3. Arruda E, Pitkaranta A, Witek TJ, Jr., Doyle CA, Hayden FG. Frequency and natural history of rhinovirus infections in adults during autumn. *J Clin Microbiol* 1997; **35**(11): 2864-8.
4. Fendrick AM, Monto AS, Nightengale B, Sarnes M. The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch Intern Med* 2003; **163**(4): 487-94.
5. Bramley TJ, Lerner D, Sarnes M. Productivity losses related to the common cold. *J Occup Environ Med* 2002; **44**(9): 822-9.
6. Jain S, Self WH, Wunderink RG, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med* 2015; **373**(5): 415-27.
7. Quan TP, Fawcett NJ, Wrightson JM, et al. Increasing burden of community-acquired pneumonia leading to hospitalisation, 1998–2014. *Thorax* 2016; **71**(6): 535.
8. Chalmers J, Campling J, Ellsbury G, Hawkey PM, Madhava H, Slack M. Community-acquired pneumonia in the United Kingdom: a call to action. *Pneumonia* 2017; **9**(1): 15.
9. Singanayagam A, Chalmers JD. Severity assessment scores to guide empirical use of antibiotics in community acquired pneumonia. *The Lancet Respiratory Medicine* 2013; **1**(8): 653-62.
10. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive Care Medicine* 2010; **36**(2): 222-31.
11. Burkhardt O, Ewig S, Haagen U, et al. Procalcitonin guidance and reduction of antibiotic use in acute respiratory tract infection. *European Respiratory Journal* 2010; **36**(3): 601.
12. Huang DT, Yealy DM, Filbin MR, et al. Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection. *New England Journal of Medicine* 2018; **379**(3): 236-49.
13. Wu J-Y, Lee S-H, Shen C-J, et al. Use of serum procalcitonin to detect bacterial infection in patients with autoimmune diseases: A systematic review and meta-analysis. *Arthritis & Rheumatism* 2012; **64**(9): 3034-42.
14. Qureshi H, Sharafkhaneh A, Hanania NA. Chronic obstructive pulmonary disease exacerbations: latest evidence and clinical implications. *Therapeutic Advances in Chronic Disease* 2014; **5**(5): 212-27.

15. Hewitt R, Farne H, Ritchie A, Luke E, Johnston SL, Mallia P. The role of viral infections in exacerbations of chronic obstructive pulmonary disease and asthma. *Therapeutic Advances in Respiratory Disease* 2015; **10**(2): 158-74.
16. <https://www.erswhitebook.org/chapters/chronic-obstructive-pulmonary-disease/>. European Lung White book. 2018 (accessed 07/06/18 2018).
17. <https://statistics.blf.org.uk/copd>. Chronic obstructive pulmonary disease (COPD) statistics. 2018 (accessed 12/05/2018 2018).
18. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**(9859): 2095-128.
19. Marsh S, Aldington S, Shirtcliffe P, Weatherall M, Beasley R. Smoking and COPD: what really are the risks? *European Respiratory Journal* 2006; **28**(4): 883.
20. Park HY, Sin D. Smoking kills, quitting heals: the importance of smoking cessation in COPD. *The Clinical Respiratory Journal* 2011; **5**(4): 185-6.
21. Gold PM. The 2007 GOLD Guidelines: A Comprehensive Care Framework. *Respiratory Care* 2009; **54**(8): 1040.
22. Sze MA, Dimitriu PA, Hayashi S, et al. The Lung Tissue Microbiome in Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* 2012; **185**(10): 1073-80.
23. Caverly LJ, Huang YJ, Sze MA. Past, Present, and Future Research on the Lung Microbiome in Inflammatory Airway Disease. *Chest* 2019; **156**(2): 376-82.
24. Sze MA, Dimitriu PA, Suzuki M, et al. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* 2015; **192**(4): 438-45.
25. Hilty M, Burke C, Pedro H, et al. Disordered Microbial Communities in Asthmatic Airways. *PLOS ONE* 2010; **5**(1): e8578.
26. Wang L, Hao K, Yang T, Wang C. Role of the Lung Microbiome in the Pathogenesis of Chronic Obstructive Pulmonary Disease. *Chin Med J (Engl)* 2017; **130**(17): 2107-11.
27. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet* 2014; **384**(9944): 691-702.
28. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; **176**(6): 532-55.
29. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine* 2010; **363**(12): 1128-38.
30. Calverley PMA. Chronic Obstructive Pulmonary Disease Exacerbations and Lung Function Decline. Mechanism or Marker? *American Journal of Respiratory and Critical Care Medicine* 2017; **195**(3): 278-9.

31. David A, Corlateanu A. Influence of COPD exacerbations on health related quality of life. *European Respiratory Journal* 2012; **40**(Suppl 56): P4821.
32. Ferrer M AJ, Morera J, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. *Annals of Internal Medicine* 1997; **127**: 1072-9.
33. Sethi S, Murphy TF. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med* 2008; **359**(22): 2355-65.
34. Wedzicha JA. Role of Viruses in Exacerbations of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society* 2004; **1**(2): 115-20.
35. Kurai D, Saraya T, Ishii H, Takizawa H. Virus-induced exacerbations in asthma and COPD. *Front Microbiol* 2013; **4**: 293.
36. Mallia P, Message SD, Gielen V, et al. Experimental rhinovirus infection as a human model of chronic obstructive pulmonary disease exacerbation. *Am J Respir Crit Care Med* 2011; **183**(6): 734-42.
37. Beasley V, Joshi PV, Singanayagam A, Molyneaux PL, Johnston SL, Mallia P. Lung microbiology and exacerbations in COPD. *International journal of chronic obstructive pulmonary disease* 2012; **7**: 555-69.
38. Rosell A, Monso E, Soler N, et al. Microbiologic determinants of exacerbation in chronic obstructive pulmonary disease. *Arch Intern Med* 2005; **165**(8): 891-7.
39. Hurst JR, Perera WR, Wilkinson TM, Donaldson GC, Wedzicha JA. Systemic and upper and lower airway inflammation at exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; **173**(1): 71-8.
40. Wilkinson TM, Hurst JR, Perera WR, Wilks M, Donaldson GC, Wedzicha JA. Effect of interactions between lower airway bacterial and rhinoviral infection in exacerbations of COPD. *Chest* 2006; **129**(2): 317-24.
41. Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med* 2002; **347**(7): 465-71.
42. Monsó E. Microbiome in chronic obstructive pulmonary disease. *Annals of translational medicine* 2017; **5**(12): 251-.
43. Wang Z, Bafadhel M, Haldar K, et al. Lung microbiome dynamics in chronic obstructive pulmonary disease exacerbations. *European Respiratory Journal* 2016: ERJ-01406-2015.
44. George SN, Garcha DS, Mackay AJ, et al. Human rhinovirus infection during naturally occurring COPD exacerbations. *Eur Respir J* 2014; **44**(1): 87-96.
45. Mallia P, Footitt J, Sotero R, et al. Rhinovirus infection induces degradation of antimicrobial peptides and secondary bacterial infection in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012; **186**(11): 1117-24.
46. Ballinger MN, Standiford TJ. Postinfluenza bacterial pneumonia: host defenses gone awry. *J Interferon Cytokine Res* 2010; **30**(9): 643-52.
47. Sun K, Metzger DW. Inhibition of pulmonary antibacterial defense by interferon-gamma during recovery from influenza infection. *Nat Med* 2008; **14**(5): 558-64.

48. Didierlaurent A, Goulding J, Patel S, et al. Sustained desensitization to bacterial Toll-like receptor ligands after resolution of respiratory influenza infection. *J Exp Med* 2008; **205**(2): 323-9.
49. Roede BM, Bresser P, Prins JM, Schellevis F, Verheij TJ, Bindels PJ. Reduced risk of next exacerbation and mortality associated with antibiotic use in COPD. *Eur Respir J* 2009; **33**(2): 282-8.
50. Aaron SD, Donaldson GC, Whitmore GA, Hurst JR, Ramsay T, Wedzicha JA. Time course and pattern of COPD exacerbation onset. *Thorax* 2012; **67**(3): 238.
51. Daniels JMA, Boersma WG. Criteria for Antibiotic Therapy in Acute Exacerbations of COPD. *American Journal of Respiratory and Critical Care Medicine* 2010; **182**(9): 1207-8.
52. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med* 1987; **106**(2): 196-204.
53. Hayden RT, Yan X, Wick MT, et al. Factors contributing to variability of quantitative viral PCR results in proficiency testing samples: a multivariate analysis. *J Clin Microbiol* 2012; **50**(2): 337-45.
54. Bafadhel M, McKenna S, Terry S, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. *Am J Respir Crit Care Med* 2011; **184**(6): 662-71.
55. Goossens H, Ferech M, Vander Stichele R, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005; **365**(9459): 579-87.
56. Shallcross LJ, Davies DSC. Antibiotic overuse: a key driver of antimicrobial resistance. *The British journal of general practice : the journal of the Royal College of General Practitioners* 2014; **64**(629): 604-5.
57. Lessa FC, Mu Y, Bamberg WM, et al. Burden of Clostridium difficile infection in the United States. *N Engl J Med* 2015; **372**(9): 825-34.
58. López-Campos JL, Hartl S, Pozo-Rodriguez F, Roberts CM, European CA. Antibiotic Prescription for COPD Exacerbations Admitted to Hospital: European COPD Audit. *PloS one* 2015; **10**(4): e0124374-e.
59. Pauwels RA BA, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *. American journal of respiratory and critical care medicine* 2001; **163**(5): 536-42.
60. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax* 2012; **67**(1): 71-9.
61. Riedel S, Beekmann S, Heilmann KP, et al. Antimicrobial use in Europe and antimicrobial resistance in Streptococcus pneumoniae; 2007.
62. Das S, Pal S, Mitra M. Significance of Exhaled Breath Test in Clinical Diagnosis: A Special Focus on the Detection of Diabetes Mellitus. *Journal of Medical and Biological Engineering* 2016; **36**(5): 605-24.

63. Mazzatenta A, Di Giulio C, Pokorski M. Pathologies currently identified by exhaled biomarkers. *Respiratory Physiology & Neurobiology* 2013; **187**(1): 128-34.
64. Buszewski B, Keszy M, Ligor T, Amann A. Human exhaled air analytics: biomarkers of diseases. *Biomed Chromatogr* 2007; **21**(6): 553-66.
65. Amann A, Costello Bde L, Miekisch W, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J Breath Res* 2014; **8**(3): 034001.
66. Shimamoto C, Hirata I, Katsu K. Breath and blood ammonia in liver cirrhosis. *Hepatogastroenterology* 2000; **47**(32): 443-5.
67. Brooks SW, Moore DR, Marzouk EB, Glenn FR, Hallock RM. Canine olfaction and electronic nose detection of volatile organic compounds in the detection of cancer: a review. *Cancer Invest* 2015; **33**(9): 411-9.
68. Minh TDC, Blake DR, Galassetti PR. The Clinical Potential of Exhaled Breath Analysis For Diabetes Mellitus. *Diabetes research and clinical practice* 2012; **97**(2): 195-205.
69. Wells DL, Lawson SW, Siriwardena AN. Canine responses to hypoglycemia in patients with type 1 diabetes. *J Altern Complement Med* 2008; **14**(10): 1235-41.
70. J Dalziel D, Uthman B, P Mcgorray S, L Reep R. Seizure-alert dogs: A review and preliminary study; 2003.
71. Kharitonov SA, Yates D, Robbins RA, Logan-Sinclair R, Shinebourne EA, Barnes PJ. Increased nitric oxide in exhaled air of asthmatic patients. *Lancet* 1994; **343**(8890): 133-5.
72. Graham DY, Klein PD, Evans DJ, Jr., et al. Campylobacter pylori detected noninvasively by the 13C-urea breath test. *Lancet* 1987; **1**(8543): 1174-7.
73. Wang XR, Cassells J, Berna AZ. Stability control for breath analysis using GC-MS. *Journal of Chromatography B* 2018; **1097-1098**: 27-34.
74. Amann A, Poupart G, Telser S, Ledochowski M, Schmid A, Mechtcheriakov S. Applications of breath gas analysis in medicine. *International Journal of Mass Spectrometry* 2004; **239**(2): 227-33.
75. Schubert JK, Geiger K. Importance and perspectives of breath analysis. *Anesthesiologie Intensivmedizin Notfallmedizin Schmerztherapie* 1999; **34**(7): 391-5.
76. Lind MV, Savolainen OI, Ross AB. The use of mass spectrometry for analysing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large numbers of biological samples. *European Journal of Epidemiology* 2016; **31**(8): 717-33.
77. Kanani H, Chrysanthopoulos PK, Klapa MI. Standardizing GC-MS metabolomics. *Journal of Chromatography B* 2008; **871**(2): 191-201.
78. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* 2012; **4**(18): 2249-64.

79. Wang H, Muehlbauer MJ, O'Neal SK, et al. Recommendations for Improving Identification and Quantification in Non-Targeted, GC-MS-Based Metabolomic Profiling of Human Plasma. *Metabolites* 2017; **7**(3): 45.
80. Khoubnasabjafari M, Rahimpour E, Jouyban A. Exhaled breath condensate as an alternative sample for drug monitoring. *Bioanalysis* 2018; **10**(2): 61-4.
81. Fens N, Zwinderman AH, van der Schee MP, et al. Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. *Am J Respir Crit Care Med* 2009; **180**(11): 1076-82.
82. Ellis AM, Mayhew CA. Proton transfer reaction mass spectrometry: principles and applications: John Wiley & Sons; 2013.
83. Ho CS, Lam CWK, Chan MHM, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical biochemist Reviews* 2003; **24**(1): 3-12.
84. Millington DS, Kodo N, Norwood DL, Roe CR. Tandem mass spectrometry: a new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism. *J Inherit Metab Dis* 1990; **13**(3): 321-4.
85. Fitzgerald RL, Herold DA. Serum total testosterone: immunoassay compared with negative chemical ionization gas chromatography-mass spectrometry. *Clinical Chemistry* 1996; **42**(5): 749.
86. Strathmann FG, Hoofnagle AN. Current and Future Applications of Mass Spectrometry to the Clinical Laboratory. *American Journal of Clinical Pathology* 2011; **136**(4): 609-16.
87. Beale DJ, Jones OAH, Karpe AV, et al. A Review of Analytical Techniques and Their Application in Disease Diagnosis in Breathomics and Salivaomics Research. *International journal of molecular sciences* 2016; **18**(1): 24.
88. Gushue JN. Chapter 11 - Principles and Applications of Gas Chromatography Quadrupole Time-of-Flight Mass Spectrometry. In: Ferrer I, Thurman EM, eds. *Comprehensive Analytical Chemistry*: Elsevier; 2013: 255-70.
89. Marriott P, Shellie R. Principles and applications of comprehensive two-dimensional gas chromatography; 2002.
90. Xue Z, Lixin D, Qi X. Gas Chromatography Mass Spectrometry Coupling Techniques; 2015.
91. Tan LT. Review of GC/MS: A Practical User's Guide, 2nd Edition. *Journal of Natural Products* 2016; **79**(10): 2763-.
92. Lawal O, Ahmed WM, Nijssen TME, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics* 2017; **13**(10): 110.
93. Sanders JK. GC/MS: A Practical User's Guide (McMaster, Marvin; McMaster, Christopher). *Journal of Chemical Education* 2000; **77**(10): 1282.
94. Fialkov AB, Steiner U, Lehotay SJ, Amirav A. Sensitivity and noise in GC-MS: Achieving low limits of detection for difficult analytes. *International Journal of Mass Spectrometry* 2007; **260**(1): 31-48.

95. Shirasu M, Touhara K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J Biochem* 2011; **150**(3): 257-66.
96. Phillips M, Cataneo RN, Condos R, et al. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis (Edinb)* 2007; **87**(1): 44-52.
97. Sethi S, Nanda R, Chakraborty T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin Microbiol Rev* 2013; **26**(3): 462-75.
98. Heddergott C, Calvo AM, Latgé JP. The Volatome of *Aspergillus fumigatus*. *Eukaryotic Cell* 2014; **13**(8): 1014-25.
99. Montuschi P, Santonico M, Mondino C, et al. Diagnostic performance of an electronic nose, fractional exhaled nitric oxide, and lung function testing in asthma. *Chest* 2010; **137**(4): 790-6.
100. Phillips CO, Syed Y, Parthala NM, Zwiggelaar R, Claypole TC, Lewis KE. Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls. *J Breath Res* 2012; **6**(3): 036003.
101. Lamote K, Brinkman P, Vandermeersch L, et al. Breath analysis by gas chromatography-mass spectrometry and electronic nose to screen for pleural mesothelioma: a cross-sectional case-control study. *Oncotarget* 2017; **8**(53): 91593-602.
102. Machado RF, Laskowski D, Deffenderfer O, et al. Detection of lung cancer by sensor array analyses of exhaled breath. *Am J Respir Crit Care Med* 2005; **171**(11): 1286-91.
103. BTS. <https://www.brit-thoracic.org.uk/media/382969/thoraxjnl-2017-211189.pdf>.
104. van de Kant KD, van der Sande LJ, Jobsis Q, van Schayck OC, Dompeling E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir Res* 2012; **13**: 117.
105. Florio W, Tavanti A, Barnini S, Ghelardi E, Lupetti A. Recent Advances and Ongoing Challenges in the Diagnosis of Microbial Infections by MALDI-TOF Mass Spectrometry. *Frontiers in Microbiology* 2018; **9**(1097).
106. Kootallur BN, Thangavelu CP, Mani M. Bacterial identification in the diagnostic laboratory: how much is enough? *Indian J Med Microbiol* 2011; **29**(4): 336-40.
107. Strejcek M, Smrhova T, Junkova P, Uhlik O. Whole-Cell MALDI-TOF MS Versus 16S rRNA Gene Analysis for Identification and Dereplication of Recurrent Bacterial Isolates. *Front Microbiol* 2018; **9**: 1294.
108. Kostrzewa M. Application of the MALDI Biotyper to clinical microbiology: progress and potential. *Expert Review of Proteomics* 2018; **15**(3): 193-202.
109. Traxler S, Bischoff A-C, Saß R, et al. VOC breath profile in spontaneously breathing awake swine during Influenza A infection. *Scientific Reports* 2018; **8**(1): 14857.
110. Schivo M, Aksenov AA, Linderholm AL, et al. Volatile emanations from in vitro airway cells infected with human rhinovirus. *J Breath Res* 2014; **8**(3): 037110.
111. Aksenov AA, Sandrock CE, Zhao W, et al. Cellular scent of influenza virus infection. *Chembiochem* 2014; **15**(7): 1040-8.

112. Schwarz KB. Oxidative stress during viral infection: a review. *Free Radic Biol Med* 1996; **21**(5): 641-9.
113. Phillips M, Cataneo RN, Greenberg J, Gunawardena R, Naidu A, Rahbari-Oskouei F. Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress. *J Lab Clin Med* 2000; **136**(3): 243-9.
114. van Geffen WH, Bruins M, Kerstjens HA. Diagnosing viral and bacterial respiratory infections in acute COPD exacerbations by an electronic nose: a pilot study. *J Breath Res* 2016; **10**(3): 036001.
115. Kumar S, Huang J, Abbassi-Ghadi N, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. *Ann Surg* 2015; **262**(6): 981-90.
116. Scott-Thomas A, Epton M, Chambers S. Validating a breath collection and analysis system for the new tuberculosis breath test. *J Breath Res* 2013; **7**(3): 037108.
117. Allardyce RA, Hill AL, Murdoch DR. The rapid evaluation of bacterial growth and antibiotic susceptibility in blood cultures by selected ion flow tube mass spectrometry. *Diagn Microbiol Infect Dis* 2006; **55**(4): 255-61.
118. Chippendale TWE, Gilchrist FJ, Španěl P, Alcock A, Lenney W, Smith D. Quantification by SIFT-MS of volatile compounds emitted by *Aspergillus fumigatus* cultures and in co-culture with *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. *Anal Methods* 2014; **6**(20): 8154-64.
119. Filipiak W, Sponring A, Baur MM, et al. Characterization of volatile metabolites taken up by or released from *Streptococcus pneumoniae* and *Haemophilus influenzae* by using GC-MS. *Microbiology* 2012; **158**(Pt 12): 3044-53.
120. Hanaa S, Federico F, Jose Luis M, et al. Using the Electronic Nose to Identify Airway Infection during COPD Exacerbations. *PLoS ONE* 2015; **10**(9): e0135199.
121. Sibila O, Garcia-Bellmunt L, Giner J, et al. Identification of airway bacterial colonization by an electronic nose in Chronic Obstructive Pulmonary Disease. *Respir Med* 2014; **108**(11): 1608-14.
122. Kubáň P, Foret F. Exhaled breath condensate: Determination of non-volatile compounds and their potential for clinical diagnosis and monitoring. A review. *Analytica Chimica Acta* 2013; **805**: 1-18.
123. Allardyce RA, Langford VS, Hill AL, Murdoch DR. Detection of volatile metabolites produced by bacterial growth in blood culture media by selected ion flow tube mass spectrometry (SIFT-MS). *J Microbiol Methods* 2006; **65**(2): 361-5.
124. Scotter JM, Allardyce RA, Langford VS, Hill A, Murdoch DR. The rapid evaluation of bacterial growth in blood cultures by selected ion flow tube-mass spectrometry (SIFT-MS) and comparison with the BacT/ALERT automated blood culture system. *Journal of Microbiological Methods* 2006; **65**(3): 628-31.

125. Filipiak W, Sponring A, Baur MM, et al. Molecular analysis of volatile metabolites released specifically by *Staphylococcus aureus* and *Pseudomonas aeruginosa*. *BMC Microbiol* 2012; **12**: 113.
126. Junger M, Vautz W, Kuhns M, et al. Ion mobility spectrometry for microbial volatile organic compounds: a new identification tool for human pathogenic bacteria. *Appl Microbiol Biotechnol* 2012; **93**(6): 2603-14.
127. Zhu J, Bean HD, Jimenez-Diaz J, Hill JE. Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study. *J Appl Physiol (1985)* 2013; **114**(11): 1544-9.
128. Abd El Qader A, Lieberman D, Shemer Avni Y, et al. Volatile organic compounds generated by cultures of bacteria and viruses associated with respiratory infections. *Biomed Chromatogr* 2015; **29**(12): 1783-90.
129. Bos LD, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. *PLoS Pathog* 2013; **9**(5): e1003311.
130. Papadimitropoulos MP, Vasilopoulou CG, Maga-Nteve C, Klapa MI. Untargeted GC-MS Metabolomics. *Methods Mol Biol* 2018; **1738**: 133-47.
131. Markar SR, Wiggins T, Antonowicz S, et al. Assessment of a noninvasive exhaled breath test for the diagnosis of oesophagogastric cancer. *JAMA Oncology* 2018.
132. Doran S, Romano A, Hanna GB. Optimization of sampling parameters for standardized exhaled breath sampling. *J Breath Res* 2017.
133. Chin S-T, Romano A, Doran SLF, Hanna GB. Cross-platform mass spectrometry annotation in breathomics of oesophageal-gastric cancer. *Scientific Reports* 2018; **8**(1): 5139.
134. Galea D, Laponogov I, Veselkov K. Chapter 10 - Data-Driven Visualizations in Metabolic Phenotyping. In: Lindon JC, Nicholson JK, Holmes E, eds. *The Handbook of Metabolic Phenotyping*: Elsevier; 2019: 309-28.
135. Monteiro M, Moreira N, Pinto J, et al. GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. *J Cell Mol Med* 2017; **21**(9): 2092-105.
136. Nizio KD, Perrault KA, Troobnikoff AN, et al. In vitro volatile organic compound profiling using GCxGC-TOFMS to differentiate bacteria associated with lung infections: a proof-of-concept study. *J Breath Res* 2016; **10**(2): 026008.
137. Karami N, Mirzajani F, Rezadoost H, et al. Initial study of three different pathogenic microorganisms by gas chromatography-mass spectrometry. *F1000Research* 2018; **6**: 1415-.
138. Schulz S, Dickschat JS. Bacterial volatiles: the smell of small organisms. *Nat Prod Rep* 2007; **24**(4): 814-42.
139. Choi YJ, Lee SY. Microbial production of short-chain alkanes. *Nature* 2013; **502**(7472): 571-4.
140. Amann A, Corradi M, Mazzone P, Mutti A. Lung cancer biomarkers in exhaled breath. *Expert Rev Mol Diagn* 2011; **11**(2): 207-17.

141. Steeghs MM, Cristescu SM, Harren FJ. The suitability of Tedlar bags for breath sampling in medical diagnostic research. *Physiol Meas* 2007; **28**(1): 73-84.
142. Mochalski P, King J, Unterkofler K, Amann A. Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags. *Analyst* 2013; **138**(5): 1405-18.
143. Mathieu E, Escribano-Vazquez U, Descamps D, et al. Paradigms of Lung Microbiota Functions in Health and Disease, Particularly, in Asthma. *Front Physiol* 2018; **9**: 1168-.
144. Muller-Decker K, Amtmann E, Sauer G. Inhibition of the phosphorylation of the regulatory non-structural protein of vesicular stomatitis virus by an antiviral xanthate compound. *J Gen Virol* 1987; **68** (Pt 12): 3045-56.
145. Ghosh S, Kim K-H, Sohn JR. Some Insights into Analytical Bias Involved in the Application of Grab Sampling for Volatile Organic Compounds: A Case Study against Used Tedlar Bags. *The Scientific World Journal* 2011; **11**: 2160-77.
146. Sanchez EL, Lagunoff M. Viral Activation of Cellular Metabolism. *Virology* 2015; **0**: 609-18.
147. Filipiak W, Mochalski P, Filipiak A, et al. A Compendium of Volatile Organic Compounds (VOCs) Released By Human Cell Lines. *Curr Med Chem* 2016; **23**(20): 2112-31.
148. Schwarz KB. Oxidative stress during viral infection: A review. *Free Radical Biology and Medicine* 1996; **21**(5): 641-9.
149. Spencer CM, Schafer XL, Moorman NJ, Munger J. Human cytomegalovirus induces the activity and expression of acetyl-coenzyme A carboxylase, a fatty acid biosynthetic enzyme whose inhibition attenuates viral replication. *J Virol* 2011; **85**(12): 5814-24.
150. Koyuncu E, Purdy JG, Rabinowitz JD, Shenk T. Saturated Very Long Chain Fatty Acids Are Required for the Production of Infectious Human Cytomegalovirus Progeny. *PLOS Pathogens* 2013; **9**(5): e1003333.
151. del Castillo-Lozano ML, Mansour S, Tâche R, Bonnarne P, Landaud S. The effect of cysteine on production of volatile sulphur compounds by cheese-ripening bacteria. *International Journal of Food Microbiology* 2008; **122**(3): 321-7.
152. Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. *Journal of hepatology* 2016; **64**(1 Suppl): S4-S16.
153. Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine* 2008; **26 Suppl 4**: D49-53.
154. Aksenov AA, Gojova A, Zhao W, et al. Characterization of volatile organic compounds in human leukocyte antigen heterologous expression systems: a cell's "chemical odor fingerprint". *Chembiochem* 2012; **13**(7): 1053-9.
155. Gern JE. The ABCs of rhinoviruses, wheezing, and asthma. *J Virol* 2010; **84**(15): 7418-26.
156. Britto CJ, Brady V, Lee S, Dela Cruz CS. Respiratory Viral Infections in Chronic Lung Diseases. *Clin Chest Med* 2017; **38**(1): 87-96.

157. Clarsund M, Fornbacke M, Uller L, Johnston SL, Emanuelsson CAM. A Randomized, Double-Blind, Placebo-Controlled Pilot Clinical Study on ColdZyme[®] Mouth Spray against Rhinovirus-Induced Common Cold. *Open Journal of Respiratory Diseases* 2017; **Vol.07No.04**: 11.
158. Vareille M, Kieninger E, Edwards MR, Regamey N. The airway epithelium: soldier in the fight against respiratory viruses. *Clin Microbiol Rev* 2011; **24**(1): 210-29.
159. Stefansson B, Gudmundsdóttir A, Clarsund M. A medical device forming a protective barrier that deactivates four major common cold viruses; 2017.
160. Jackson DJ, Glanville N, Trujillo-Torralbo MB, et al. Interleukin-18 is associated with protection against rhinovirus-induced colds and asthma exacerbations. *Clin Infect Dis* 2015; **60**(10): 1528-31.
161. Owen AB, Perry PO. Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Ann Appl Stat* 2009; **3**(2): 564-94.
162. Veselkov KA, Mirnezami R, Strittmatter N, et al. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc Natl Acad Sci U S A* 2014; **111**(3): 1216-21.
163. Moye L. Statistical Methods for Cardiovascular Researchers. *Circ Res* 2016; **118**(3): 439-53.
164. Benjamini Y. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72**(4): 405-16.
165. Costello BdL, Amann A, Al-Kateb H, et al. A review of the volatiles from the healthy human body. *Journal of Breath Research* 2014; **8**(1): 014001.
166. Choi KH. Viral polymerases. *Advances in experimental medicine and biology* 2012; **726**: 267-304.
167. Macintyre S. Gender differences in the perceptions of common cold symptoms. *Soc Sci Med* 1993; **36**(1): 15-20.
168. Wedzicha JA, Singh R, Mackay AJ. Acute COPD exacerbations. *Clin Chest Med* 2014; **35**(1): 157-63.
169. Romano A, Hanna GB. Identification and quantification of VOCs by proton transfer reaction time of flight mass spectrometry: An experimental workflow for the optimization of specificity, sensitivity, and accuracy. *Journal of Mass Spectrometry* 2018; **53**(4): 287-95.
170. Capone S, Tufariello M, Forleo A, et al. Chromatographic analysis of VOC patterns in exhaled breath from smokers and nonsmokers. *Biomed Chromatogr* 2018; **32**(4).
171. Pizzini A, Filipiak W, Wille J, et al. Analysis of volatile organic compounds in the breath of patients with stable or acute exacerbation of chronic obstructive pulmonary disease. *Journal of Breath Research* 2018; **12**(3): 036002.
172. Singh S, Brocker C, Koppaka V, et al. Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. *Free Radic Biol Med* 2013; **56**: 89-101.

173. Fuchs P, Loeseken C, Schubert JK, Miekisch W. Breath gas aldehydes as biomarkers of lung cancer. *Int J Cancer* 2010; **126**(11): 2663-70.
174. Jelski W, Szmitkowski M. Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) in the cancer diseases. *Clin Chim Acta* 2008; **395**(1-2): 1-5.
175. Yokoyama A, Watanabe H, Fukuda H, et al. Multiple cancers associated with esophageal and oropharyngolaryngeal squamous cell carcinoma and the aldehyde dehydrogenase-2 genotype in male Japanese drinkers. *Cancer Epidemiol Biomarkers Prev* 2002; **11**(9): 895-900.
176. Schwabe RF, Jobin C. The microbiome and cancer. *Nature reviews Cancer* 2013; **13**(11): 800-12.
177. Tsuruya A, Kuwahara A, Saito Y, et al. Major Anaerobic Bacteria Responsible for the Production of Carcinogenic Acetaldehyde from Ethanol in the Colon and Rectum. *Alcohol Alcohol* 2016; **51**(4): 395-401.
178. Ackermann WW, With the Technical Assistance of Elaine K, Hilda K. THE RELATION OF THE KREBS CYCLE TO VIRAL SYNTHESIS : II. THE EFFECT OF SODIUM FLUOROACETATE ON THE PROPAGATION OF INFLUENZA VIRUS IN MICE. *The Journal of Experimental Medicine* 1951; **93**(6): 635-42.
179. Crosta P. Viruses: What are they and what they do? 2017. <https://www.medicalnewstoday.com/articles/158179.php30> May. 2017).
180. Hayden FG, Monto AS. Oral rimantadine hydrochloride therapy of influenza A virus H3N2 subtype infection in adults. *Antimicrob Agents Chemother* 1986; **29**(2): 339-41.
181. Husson M-O, Ley D, Portal C, et al. Modulation of host defence against bacterial and viral infections by omega-3 polyunsaturated fatty acids. *Journal of Infection* 2016; **73**(6): 523-35.
182. Mittal M, Siddiqui MR, Tran K, Reddy SP, Malik AB. Reactive oxygen species in inflammation and tissue injury. *Antioxid Redox Signal* 2014; **20**(7): 1126-67.
183. Yu L, Gan X, Zhou D, He F, Zeng S, Hu D. Synthesis and Antiviral Activity of Novel 1,4-Pentadien-3-one Derivatives Containing a 1,3,4-Thiadiazole Moiety. *Molecules* 2017; **22**(4).
184. Rice M, Harris G. Rice ME, Harris GT Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum Behav* 29: 615-620; 2005.
185. Johnston SL, Sanderson G, Pattemore PK, et al. Use of polymerase chain reaction for diagnosis of picornavirus infection in subjects with and without respiratory symptoms. *J Clin Microbiol* 1993; **31**(1): 111-7.
186. Papadopoulos NG, Hunter J, Sanderson G, Meyer J, Johnston SL. Rhinovirus identification by BglII digestion of picornavirus RT-PCR amplicons. *J Virol Methods* 1999; **80**(2): 179-85.
187. O'Donnell DR, McGarvey MJ, Tully JM, Balfour-Lynn IM, Openshaw PJ. Respiratory syncytial virus RNA in cells from the peripheral blood during acute infection. *J Pediatr* 1998; **133**(2): 272-4.
188. Myint S, Johnston S, Sanderson G, Simpson H. Evaluation of nested polymerase chain methods for the detection of human coronaviruses 229E and OC43. *Mol Cell Probes* 1994; **8**(5): 357-64.

189. Ellis JS, Fleming DM, Zambon MC. Multiplex reverse transcription-PCR for surveillance of influenza A and B viruses in England and Wales in 1995 and 1996. *J Clin Microbiol* 1997; **35**(8): 2076-82.
190. Corne JM, Green S, Sanderson G, Caul EO, Johnston SL. A multiplex RT-PCR for the detection of parainfluenza viruses 1-3 in clinical samples. *J Virol Methods* 1999; **82**(1): 9-18.
191. Hierholzer JC, Halonen PE, Dahlen PO, Bingham PG, McDonough MM. Detection of adenovirus in clinical specimens by polymerase chain reaction and liquid-phase hybridization quantitated by time-resolved fluorometry. *J Clin Microbiol* 1993; **31**(7): 1886-91.
192. Cunningham AF, Johnston SL, Julious SA, Lampe FC, Ward ME. Chronic Chlamydia pneumoniae infection and asthma exacerbations in children. *Eur Respir J* 1998; **11**(2): 345-9.
193. van Kuppeveld FJ, Johansson KE, Galama JM, et al. 16S rRNA based polymerase chain reaction compared with culture and serological methods for diagnosis of Mycoplasma pneumoniae infection. *Eur J Clin Microbiol Infect Dis* 1994; **13**(5): 401-5.