

ASSESSING THE MICROBIOME MEDIATED AUGMENTED VOLATILE ORGANIC COMPOUNDS RESPONSE IN OESOPHAGOGASTRIC CANCER

Bhamini Vadhvana MBChB BSc (Hons) MRCS

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

2022

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

Supervisors

Professor George B. Hanna FRCS PhD

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

Mr Piers R. Boshier MRCS PhD

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

DECLARATION OF ORIGINALITY

I hereby declare that I am the sole author of this Thesis and that all the work within it is my own. Any individuals who carried out collaborative work are appropriately credited.



Bhamini Vadhvana

25th April 2022

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

ABSTRACT

Oesophagogastric cancer is the fifth most common cancer in England and Wales. Early non-specific symptoms are shared with benign disease presenting a diagnostic dilemma. There remains a clinical need to develop an acceptable, accurate and affordable method to identify high risk patients at an earlier stage.

The aim of this thesis is to investigate the role of augmenting target VOCs by exploiting the oral and tumour associated microbiome for accurate detection of early oesophagogastric cancer.

Microbial profiling with 16S rRNA gene sequencing (V3-V4) of oesophagogastric cancer tissue demonstrated *Lactobacillus* ($p=0.04$) and *Pseudopropionibacterium* ($p=0.047$) were significantly more abundant in gastric cancer.

Cultivated bacteria from matched tumour and saliva samples underwent Shotgun metagenomic analysis. Liquid cultures were established to monitor cell growth curves. Glucose, 0.1M concentration, was spiked into the culture and headspace VOCs extracted at the end log phase for two-dimensional gas-chromotography-mass spectrometry analysis. Medium chain fatty acids were increased following glucose introduction; (i) eicasonic acid 2.8-fold (ii) butanoic acid, 4-hydroxy 4.1-fold (iii) pentanoic acid, 3-methyl 10.7-fold. Predicted metabolic pathways demonstrated the potential role of multiple pathways in target end-metabolite production: succinate to propionate, pyruvate to acetate, glutamate to butyrate.

A clinical study was designed for patients to consume nutritional stimulant drinks and provide serial exhaled breath samples into the selected ion flow tube-mass spectrometer. Short chain fatty acids; butyric- and propanoic acids, were increased 10.5- and 4-fold respectively in gastric cancer. Butyric acid, propanoic acid and butanal were increased in both cancer types after glycerol consumption. A combination of amino acids demonstrated an increase in p-cresol.

This study supports the role of an augmented VOC breath test by exploiting the oral and tumour-associated microbiome. It offers a potentially novel non-invasive method to triage patients with non-specific upper gastrointestinal patients, facilitate earlier cancer detection and ultimately improve overall survival.

PEER REVIEWED PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

Vadhwana B, Belluomo I, Boshier PR, Pavlou C, Španěl P, Hanna GB. Impact of oral cleansing strategies on exhaled volatile organic compound levels. *Rapid Commun Mass Spectrom.* 2020;34(9):e8706

Belluomo I, Boshier PR, Myridakis A, **Vadhwana B**, Markar SR, Španěl P, Hanna GB. Selected ion flow tube mass spectrometry for targeted analysis of volatile organic compounds in human breath. *Nat Protoc.* 2021;16(7):3419-3438

PRESENTATIONS

Vadhwana B, Belluomo I, Boshier PR, Pavlou C, Španěl P, Hanna GB. Impact of oral cleansing strategies on exhaled volatile organic compound levels.
Presented and Awarded 'Best Oral Poster' presentation at the Annual Clinical Mass Spectrometry Conference 2019.

GRANTS

Innovate UK: Innovation Scholars secondments: biomedical sciences. Investigating the mechanistic pathways of augmented volatile organic compounds related to the oesophagogastric cancer microbiome.

ACKNOWLEDGEMENTS

I am eternally grateful to my supervisors **Professor George Hanna** and **Mr Piers Boshier** for the support, inspirational guidance, and invaluable mentorship throughout my research and beyond.

I am sincerely grateful for the support the following people have provided me during my research:

Dr Ilaria Belluomo; assistant supervisor – for guidance with the SIFT-MS and PTR-MS studies, and continual support throughout my PhD studies

Professor Lesley Hoyles – for performing and teaching microbial metagenomic analysis

Professor Patrik Španěl – for expert guidance with analytical strategies with SIFT-MS and GC-MS studies

Ms Sara Jamel – for assistance and guidance for all GC-MS studies including optimisation, quality control measures, sample processing and data analysis

Miss Katerina Vanessa Savva – for sampling assistance with the clinical intraoperative and developmental studies

Dr Ian Fotheringham, Ingenza Ltd – for facilitating the *in vitro* component of the augmentation study

Ms Marianne Keith, Ingenza Ltd – for guidance with *in vitro* bacteria culturing

Dr Alison Arnold, Ingenza Ltd – for guidance with *in vitro* liquid bacteria culturing

Dr Fraser Brown, Ingenza Ltd – for guidance with establishing a system for headspace VOC sampling

Dr Claire Wilson, Ingenza Ltd – for guidance with developing a biotransformation workflow for bacteria headspace VOC sampling

Dr Nathan Danckert – for guidance and assistance with 16S rRNA gene microbial analysis

Dr Aaron Parker, SepSolve Analytical – for supporting and performing headspace VOC analysis of saliva using the Centri® 1D GC-MS

Mr Christopher Peters, clinical UGI and anaesthetic team – for their support with clinical sampling

Mr Jack James (medical student, BSc) – for assistance with the saliva methodology work

Miss Melina Pelling (medical student, UROP) – for assistance with saliva VOC variation work

Table of Contents

DECLARATION OF ORIGINALITY	2
COPYRIGHT DECLARATION	2
ABSTRACT	3
PEER REVIEWED PUBLICATIONS AND PRESENTATIONS	4
ACKNOWLEDGEMENTS	5
LIST OF FIGURES	8
LIST OF TABLES	11
ABBREVIATIONS	13
SECTION 1	14
LITERATURE REVIEW	14
1. CHAPTER 1: EARLY DETECTION OF OESOPHAGOGASTRIC CANCER: AN UNMET CLINICAL NEED	15
1.1 OESOPHAGOGASTRIC CANCER	15
1.2 DIAGNOSTIC BIOMARKERS AND CHALLENGES IN BREATH ANALYSIS	16
1.3 MECHANISMS OF VOC AUGMENTATION: OESOPHAGOGASOGASTRIC CANCER MICROBIOME AND CANCER CELLS	19
1.3.1 <i>Oesophagogastric Cancer Microbiome</i>	19
1.3.2 <i>Bacterial augmentation of Volatile Organic Compounds</i>	20
1.4 THESIS HYPOTHESIS AND AIMS	21
2. CHAPTER 2: EVALUATION OF THE OESOPHAGOGASTRIC CANCER ASSOCIATED MICROBIOME: A SYSTEMATIC REVIEW AND QUALITY ASSESSMENT	22
3. ... CHAPTER 3: SYSTEMATIC REVIEW: AUGMENTATION OF CHORT CHAIN FATTY ACIDS BY BACTERIAL METABOLISM OF NUTRITIONAL METABOLIC SUBSTRATES	37
SECTION 2	48
CHARACTERISATION OF THE OESOPHAGOGASTRIC CANCER ASSOCIATED MICROBIOME	48
4. CHAPTER 4: CHARACTERISATION OF OESOPHAGOGASTRIC CANCER MICROBIOME USING 16S RNA SEQUENCING	49
5. CHAPTER 5: BACTERIA CELL CULTURE FROM OESOPHAGOGASTRIC CANCER TISSUE AND SALIVA	67
5.1 BACTERIA CELL CULTURES FROM OESOPHAGOGASTRIC CACNER TISSUE AND SALIVA USING SHOTGUN METAGENOMIC ANALYSIS	67
5.2 PRE-CULTURE 16S rRNA SEQUENCING ANALYSIS TO ASSESS CULTURE BIAS	85
SECTION 3	91
ASSESSMENT OF VOLATILE ORGANIC COMPOUNDS ASSOCIATED WITH THE ORAL MICROBIOME IN OESOPHAGOGASTRIC CANCER	91
6. ... CHAPTER 6: ASSESSING THE IMPACT OF ORAL CLEANSING STRATEGIES ON EXHALED VOLATILE ORGANIC COMPOUNDS	92
7. CHAPTER 7: METHODOLOGICAL OPTIMISATION OF SALIVARY HEADSPACE VOC ANALYSIS	102
7.1 METHODOLOGICAL OPTIMISATION OF SALIVARY HEADSPACE VOC ANALYSIS .	102

7.2 ASSESSMENT OF THE LONGITUDINAL VARIATION IN VOC LEVELS IN WHOLE SALIVA.....	109
8. CHAPTER 8: ASSESSMENT OF SALIVARY HEADSPACE VOLATILE ORGANIC COMPOUNDS IN OESOPHAGOGASTRIC CANCER.....	115
8.1 ASSESSMENT OF SALIVARY VOCs FOR THE DETECTION OF OESOPHAGOGASTRIC CANCER USING TARGETED SIFT-MS.....	115
8.2 ASSESSMENT OF SALIVARY VOCs FOR THE DETECTION OF OESOPHAGOGASTRIC CANCER USING UNTARGETED 1D GC-MS.....	119
SECTION 4.....	125
AUGMENTATION OF VOLATILE ORGANIC COMPOUNDS IN OESOPHAGOGASTRIC CANCER	125
9. CHAPTER 9: IN-VITRO AUGMENTATION OF VOC FROM OESOPHAGOGASTRIC CANCER BACTERIA CELL CULTURES.....	126
10. CHAPTER 10: AUGMENTATION OF VOCs BY ANATOMICAL COMPARTMENTALISATION OF THE UPPER GASTROINTESTINAL TRACT AND THE LUNGS	147
11. CHAPTER 11: CLINICAL VOC AUGMENTATION WITHIN EXHALED BREATH USING NUTRITIONAL METABOLIC SUBSTRATES IN OESOPHAGOGASTRIC CANCER PATIENTS ...	158
SECTION 5.....	173
FUTURE WORK	173
12 CHAPTER 12: DISCUSSION (SUMMARY, CONCLUSIONS, FUTURE WORK)	174
REFERENCES.....	178

LIST OF FIGURES

Figure 1. Flowchart describing article selection for the assessment of the oesophagogastric cancer microbiome.	25
Figure 2. Bacteria enriched in gastric adenocarcinoma, benign disease and healthy controls. Bacteria identified to be significant in 3 or more studies out of 31 studies were included. P = phyla, C = class, O = order, F= family, G = genus, Sp = species. Five genera, <i>Lactobacillus</i> , <i>Streptococcus</i> , <i>Prevotella</i> , <i>Fusobacterium</i> and <i>Veillonella</i> were demonstrated in higher abundance in gastric adenocarcinoma. 30	
Figure 3. D1000 TapeStation trace demonstrating 16S rRNA gene amplicon PCR band present at ~550bp for V3-V4 regions. EL1 (L) = ladder. A1 = blank. B1= blank. C1 = Sample 1. D1 = Sample 2. 52	
Figure 4. Alpha diversity metrics of bacteria genera comparing Control normal stomach (red) to Gastric cancer adjacent normal stomach (green) and Gastric cancer (blue); $p > 0.05$. No significant differences were identified.	54
Figure 5. Beta diversity metrics of bacteria genera comparing Control normal stomach (red) to Gastric cancer adjacent normal stomach (green) and Gastric cancer (blue); Bray-curtis $p = 0.004$, Jenson-shannon $p = 0.03$, Jaccard index $p = 0.001$, Unweighted Unifrac $p = 0.02$, Weight Unifrac $p = 0.02$	55
Figure 6. Box plots demonstrating select bacteria in higher abundances in gastric cancer tissue compared to healthy control normal stomach. P value < 0.05 is considered statistically significant. ...	57
Figure 7. Linear discriminant analysis Effect Size demonstrating bacterial genera and species abundant in gastric cancer (red) compared to healthy controls.	59
Figure 8. Box plots demonstrating select bacteria enriched in oesophageal cancer tissue compared to healthy control normal oesophagus. P value < 0.05 is considered statistically significant.	62
Figure 9. Linear discriminant analysis Effect Size demonstrating bacterial genera and species abundant in oesophageal cancer (red) compared to healthy controls.	64
Figure 10. Anaerobic and aerobic saliva growth with serial dilutions from 10^{-4} to 10^{-7} (50 μ L in 450 μ L sterile PBS) to separate colonies on fastidious anaerobic agar and nutrient agar, respectively. Sample 189 – OG56. Slower growing colonies were observed at 48 hours. No macroscopic difference was seen with a further 2 days of incubation, totalling 4 days.	69
Figure 11. Macroscopic assessment of bacteria-glycerol stocks demonstrating a similar growth pattern to the original culture, following growth on fastidious anaerobic agar. Sample 33 OG47 NO.	70
Figure 12. Number and quality of MAGs obtained from metagenomes generated from mixed bacterial cultivation work.	73
Figure 13. Phylum-level breakdown of the medium and high-quality MAGs generated from original cultivation work. n= 170 medium quality and n=118 high quality according to Bowers et al.	73
Figure 14. Phylogenetic tree, generated using PhyloPhlAn, showing the taxonomic affiliations of the medium- and high-quality MAGs recovered from 65/68 of the metagenomes derived from the original isolation work. Solid colours (corresponding to the legend) represent the phyla. Genera are labelled around the outside of the phylogenetic tree. Scale bar, normalised fraction of total branch length.	74
Figure 15. Genus-level breakdown of medium- and high-quality MAGs detected in metagenomes from cultivation work.	75
Figure 16. Primary metabolic pathways predicted to be represented across 274 medium- and high-quality MAGs for which gutSMASH returned results.	76
Figure 17. Chromatograms demonstrating good quality peaks representative of pure cultures. All three bacteria cultures were established directly from gastrooesophageal junctional cancer tissue.	78
Figure 18. Phylogenetic tree showing the relationships of the three isolates' genomes with 113 related taxa. Clades of some multi-species genera have been collapsed to aid visualisation. The tree was generated using PhyloPhlAn v0.99 and visualized using iTOL. Scale bar, normalised fraction of total branch length.	79
Figure 19. Box plots demonstrating select bacteria enriched and depleted in oesophageal cancer saliva compared to healthy control saliva. P value < 0.05 was considered statistically significant.	88
Figure 20. Variation in volatile organic compounds in response to oral cleansing measures. There is minimal variation in acetone and isoprene suggesting their systemic origin. Levels of butyric acid and ethanol reduce in response to oral cleansing. Heptanal and menthone are a source of contamination from oral cleansing ingredients.	99
Figure 21. Four methodological parameters of saliva headspace optimised by VOC concentration (ppbv) assessment: saliva collection method, saliva volume, saliva dilution, saliva acidification.	107

Figure 22. Longitudinal variation of target short chain fatty acids and aldehydes in human saliva. S6 = participant 6, S2 = participant 2. There is an acceptable level of variation of volatile compounds over the five time points assessed.	112
Figure 23. Overall longitudinal variation of the concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva, as a cohort. There is an acceptable variation in concentration of select compounds across the cohort.	113
Figure 24. Box plot showing levels of butyric acid detected within whole saliva are significantly higher in the gastric and oesophageal adenocarcinoma groups compared to healthy controls. * $p < 0.05$, ** $p < 0.01$	117
Figure 25. Box plot demonstrating the peak area of propanoic acid – ethyl ester in the headspace of saliva of oesophagogastric cancer and healthy control patients. Statistical significance was demonstrated with elevated concentrations of propanoic acid – ethyl ester in gastric cancer ($p = 0.04$) compared to healthy controls.	122
Figure 26. Box plot demonstrating the peak area of a branched butyric acid (-2-cyano-3-methyl-, ethyl ester) in the headspace of saliva of oesophagogastric cancer and healthy control patients. Median peak area calculation of gastric adenocarcinoma (median peak area = 4,292,933) is markedly higher than healthy controls (median peak area = 1,919,955), albeit not significant.	123
Figure 27. 250mL flasks for culture growth and glucose augmentation set-up, Stopcock adapters: Tap (blue) to release the seal to allow headspace extraction via the outlet (red). Round bottom flasks used for anaerobic cultures, and baffled flasks for aerobic cultures.	128
Figure 28. Cell growth curves of mixed and pure bacteria cell cultures in liquid media. Growth assays depict 4 phases of the cell cycle: (i) lag phase (ii) log/exponential phase (iii) stationary phase and (iv) death phase, for anaerobic and aerobic mixed cultures.	129
Figure 29. Biotransformation workflow describing the VOC augmentation from liquid bacteria cell cultures.	130
Figure 30. Chromatogram demonstrating augmentation of Eicosanoic acid following glucose stimulation of bacteria cell culture from saliva cultivated anaerobically.	142
Figure 31. Chromatogram demonstrating the peak area of 4-hydroxy butyric acid detected from the headspace of <i>Escherichia coli</i> stimulated with glucose.	143
Figure 32. Sankey plot demonstrating an association between 3 pure cultures, predicted metabolic pathways and end-metabolic products; specific to oesophagogastric cancer.	144
Figure 33. Fold change demonstrating the increase of hexanoic- and pentanoic acid in the endoluminal compartment following glucose wash directly over the tumour, using direct real-time analysis on the SIFT-MS.	150
Figure 34. Chromatograms of Nonanoic acid (medium chain fatty acid) demonstrating an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).	153
Figure 35. Chromatograms of n-decanoic acid (medium chain fatty acid) demonstrating an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time intervals (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).	154
Figure 36. Chromatograms of Dodecanoic acid (medium chain fatty acid) demonstrating upregulation an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time intervals (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).	155
Figure 37. Glucose consumption: Fold increase of target short chain fatty acids (acetic-, butyric- and propanoic acids) and aldehydes (butanal, pentanal and propanal) between three cohorts; oesophageal adenocarcinoma, gastric adenocarcinoma and healthy controls.	163
Figure 38. Glycerol consumption: Fold increase of target short chain fatty acids (acetic-, butyric- and propanoic acids) and aldehydes (butanal, pentanal and propanal) between three cohorts; oesophageal adenocarcinoma, gastric adenocarcinoma and healthy controls.	165
Figure 39. Combined amino acids consumption: Fold increase of p-cresol between two cohorts; oesophagogastric adenocarcinoma and healthy controls.	166
Figure 40. Potential metabolic pathway for glycerol metabolism; (i) Dehydrogenase pathway (ii) Phosphorylation pathway. Adapted from [231-233]	170
Figure 41. Potential metabolic pathway for tyrosine metabolism to (i) phenol and (ii) cresol. TPL = tyrosine phenol lyase, ThiH = tyrosine lyase, TyrB = tyrosine aminotransferase. Adapted from [234]	171

Figure 42. Potential metabolic pathway for glucose and glycerol metabolism; glycolytic pathway.
Adapted from [235-237] 172

LIST OF TABLES

Table 1. ‘Red Flag’ symptoms for suspected oesophageal and gastric cancer established by the National Institute of Clinical Excellence.	16
Table 2. Study characteristics of articles reporting on the microbiome of gastric adenocarcinoma compared to a non-cancer group.	28
Table 3. Study characteristics of articles reporting on the microbiome of oesophageal cancer compared to a non-cancer group.	32
Table 4. Specific considerations for designing a microbiome assessment study in upper gastrointestinal disease.	36
Table 5. Short chain fatty acid augmentation by bacterial metabolism using nutritional stimuli.	40
Table 6. 16S rRNA gene amplicon PCR primers.	51
Table 7. Alpha and beta diversities of bacterial communities in gastric cancer tissue, adjacent normal stomach, and normal stomach from healthy controls. NS = normal stomach. P value < 0.05 is considered statistically significant (highlighted in red).	56
Table 8. Bacteria significantly altered in gastric cancer-specific bacteria. # = high abundance. \$ = low abundance.	58
Table 9. Alpha and beta diversities of bacterial communities in oesophageal cancer tissue, adjacent normal oesophagus, and normal oesophagus from healthy controls. NO = normal oesophagus. P value < 0.05 is considered statistically significant.	61
Table 10. Bacteria significantly altered in oesophageal cancer-specific bacteria. # = high abundance. \$ = low abundance.	63
Table 11. Summary statistics for the draft genome sequences generated in this study.	77
Table 12. Alpha and beta diversity metrics (p value) for cancer bacteria samples compared to healthy controls.	87
Table 13. Details of interventions.	95
Table 14. Details of volatile organic compound analysed by selected ion flow tube mass spectrometry.	95
Table 15. Variation in selected volatile organic compound concentrations (ppbv) in response to oral cleansing measures.	98
Table 16. Concentrations of VOC of interest in control experiments assessing the VOC contribution of saliva collection equipment.	106
Table 17. Osmolality (mOsm) and flow rates (mL/min) of saliva retrieved with three different collection methods.	106
Table 18. Volatile short chain fatty acid concentrations (ppbv) to assess the optimal analysis of saliva headspace by using various methodological parameters.	106
Table 19. Longitudinal variation of the concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva, as a cohort. Samples: 1 = day 1, 2 = day 2, 3 = day 3, 4 = day 10, 5 = day 17. Coefficient of variation percentage, CV%.	111
Table 20. Baseline characteristics of patients between the three groups.	116
Table 21. Median concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva detected using the SIFT-MS (VoiceUltra200).	117
Table 22. Baseline characteristics of patients between the three groups.	120
Table 23. Optical density (OD600) measurements for liquid bacteria cultures.	133
Table 24. Mixed bacteria cultures identified by Shotgun metagenomic analysis for in vitro augmentation experimental set-up.	134
Table 25. Summary of predicted metabolic pathways associating cancer bacteria with VOCs.	137
Table 26. Volatile organic compounds increased following glucose stimulation of bacteria cell cultures derived from tumour and saliva from oesophagogastric cancer patients.	139
Table 27. Fatty acids augmented following glucose stimulation of bacteria cell cultures derived from tumour and saliva obtained from oesophagogastric cancer patients. Peak areas outputs provided by the 2D GC-MS was translated to fold change to determine augmentation.	140
Table 28. Patient and clinical demographics for the intraoperative augmentation study compartmentalising two anatomical regions (endobronchial and endoluminal). GOJ = gastroesophageal junction.	149
Table 29. Increase in volatile organic compounds in two anatomical compartments (i) endobronchial and (ii) endoluminal, following glucose wash directly over the tumour site.	152

Table 30. Nutrient stimulants selected to increase target VOCs.	159
Table 31. Baseline characteristics between the three patient groups.	160
Table 32. Glucose consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.	162
Table 33. Glycerol consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.	164
Table 34. Combined amino acids consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls and oesophagogastric adenocarcinoma.	167
Table 35. Summary of bacteria genera identified consistently between literature and original microbial analysis in gastric adenocarcinoma.	174

ABBREVIATIONS

1D GC-MS: one-dimensional Gas Chromatography – Mass Spectrometry
2D GC-MS: two-dimensional Gas Chromatography – Mass Spectrometry
ANI: average nucleotide identity
ATCC: American Type Culture Collection
AUC: area under curve
CV: coefficient of variation
DNA: deoxyribonucleic acid
dsDNA: double stranded deoxyribonucleic acid
EUT: ethanolamine utilisation
FAA: fastidious anaerobic agar
GC-MS: Gas Chromatography – Mass Spectrometry
GOJ: gastrooesophageal junction
HER2: Human Epidermal growth factor Receptor 2
IQR: interquartile range
m/z: mass-to-charge ratio
MAGs: metagenome assembled genomes
NCBI: National Center for Biotechnology Information
NCIMB: National Collection of Industrial Food and Marine Bacteria
OAC: oesophageal adenocarcinoma
OD: optical density
PBS: phosphate buffer solution
PCR: polymerase chain reaction
PD-L1: Programmed Death Ligand 1
PDU: propanediol utilisation
PFOR II: pyruvate ferredoxin oxidoreductase II
ppbv: parts per billion by volume
PTFE: polytetrafluoroethylene
PTR-ToF-MS: Proton Transfer Reaction-Time of Flight – Mass Spectrometry
PubMLST: public databases or molecular typing and microbial genome diversity)
q-PCR: quantitative polymerase chain reaction
ROC: Receiver Operating Characteristic curve
rRNA: ribosomal ribonucleic acid
RT-PCR: real time polymerase chain reaction
SCC: Squamous Cell Carcinoma
SCFA: short chain fatty acids
SIFT-MS: Selected Ion Flow Tube – Mass Spectrometry
STORMS: Strengthening The Organising and Reporting of Microbiome Studies
TCA: tricarboxylic acid
TD: thermal desorption
ToF: Time of Flight
VOC: Volatile Organic Compound

SECTION 1

Literature review

1. CHAPTER 1: EARLY DETECTION OF OESOPHAGOGASTRIC CANCER: AN UNMET CLINICAL NEED

1.1 OESOPHAGOGASTRIC CANCER

Cancer is the second leading cause of death globally, contributing to 9.6 million deaths worldwide in 2018.[1] Although cancer incidence is increasing, recorded mortality rates are declining largely due to the introduction of screening programmes, improved treatment modalities, and cancer awareness in the general population.

Oesophagogastric cancer is the fifth most common cancer in the UK with 16,000 new cases diagnosed annually and overall 5 year survival of 15%.[2] Oesophageal cancer is divided into two histological subtypes; squamous cell carcinomas (SCC) being the most common global subtype, but adenocarcinomas predominating in developed countries. Adenocarcinomas have increased in incidence in affluent countries at a rate of 6% in the last 5 years in the UK.[3] Growing obesity and lower smoking rates may be correlated with the rising incidence of adenocarcinoma in developed countries.[4]

The majority of gastric cancers are adenocarcinomas (95%). Adenocarcinomas of the cardia follow the clinical course and prognosis of oesophageal adenocarcinomas. Cardia gastric cancers are increasing in the Western world, which may be explained by their histological relationship to oesophageal adenocarcinomas.[5]

The National Oesophagogastric Cancer Audit reported only 30% of patients were offered treatment with curative intent at the time of diagnosis, with only two thirds of this cohort proceeding to surgical resection.[3] Overall, greater than 80% of patients are diagnosed with stage III and IV disease, approximately half of whom have distant metastases.[6] Critically, 5-year survival can exceed 90% in patients who are diagnosed and treated with stage I oesophageal adenocarcinoma.[7]

Poor survival from oesophagogastric cancer is principally related to its initial association with vague symptoms that leads to a delay in diagnosis. When characteristic 'Red Flag' symptoms do occur, the disease is often in an advanced, incurable stage. Table 1 details the 'Red Flag' symptoms for oesophageal and gastric cancer established by the National Institute of Clinical Excellence. Improving outcomes is dependent on either earlier detection of symptomatic patients or screening of asymptomatic high-risk patients.

Table 1. ‘Red Flag’ symptoms for suspected oesophageal and gastric cancer established by the National Institute of Clinical Excellence.

‘Red Flag’ symptoms for assessment of suspected Oesophagogastric cancer
Dysphagia, or Aged 55 years or older with weight loss and any of the following: • Upper abdominal pain • Reflux • Dyspepsia

Studies have revealed ‘red flag’ symptoms of dysphagia and weight loss have a low predictive value and cancer is diagnosed in only 1% of those referred as per the guidelines for suspected oesophagogastric cancer.[8, 9] The relatively low incidence of oesophagogastric cancer and associated costs of endoscopy (£391 per test) does not justify a screening investigation within the UK.

Non-invasive screening investigations for other cancer types have been established in clinical practice, including the faecal occult blood test, and most recently the faecal immunohistochemistry test for early detection of bowel cancer. There remains therefore an unmet clinical need for improved early detection of oesophagogastric cancer for access to curative treatment translating to significant survival benefits.

1.2 DIAGNOSTIC BIOMARKERS AND CHALLENGES IN BREATH ANALYSIS

Diagnostic biomarkers offer a non-invasive method of detection and surveillance of oesophagogastric cancer, with the potential of improving oncological outcomes. Genomic molecular targets form the majority of biomarkers investigated in tissues and various biological matrices including blood and urine. Tissue biomarkers are based on genetic and epigenetic alterations presenting a promising platform for cancer diagnosis. Molecular targets are already established in clinical practice guiding immunotherapy with HER2 expression linked with trastuzumab, and PD-L1 expression with pembrolizumab.[10, 11]

Blood biomarkers have been studied based on genomic mutations, particularly hypermethylation of target genes. Circulating free DNA is detectable in higher concentrations in cancer patients owing to additional tumour burden but is not used in clinical practice due to the potential for inflammation, trauma and exercise to influence results .[12-14] Gene dysregulation in oesophageal adenocarcinoma includes TP53 (72%), chromatin-remodelling factors with ARID1A, SMARCA4 and ARID2 (20%), and EML01 and DOCK2 (17%), and p16/CDKN2A (12%).[15] A molecular signature of six serum autoantibodies (p53, NY-ESO-1, MMP-7, HSP70, PRDX6 and BMI-1) demonstrated a sensitivity of 61.4%, specificity of 90% and AUC of 81.5% for the diagnostic accuracy in early oesophagogastric junctional adenocarcinoma.[16] Similarly, biomarkers such as TP53, E-cadherin CDH1, FGFR, PI3K, MET, VEGFR, and PD-1 have targeted cell cycle regulation, and hypermethylation of E-cadherin CDH1, p16,

RASSF1A, RUNX3 and TFP12 have been associated with gastric adenocarcinoma.[17-22] Inflammatory biomarkers have produced conflicting results to date.[23-27] Use of combined biomarkers remain in the discovery phase although PG1 and PG2, gastrin-17 and *Helicobacter pylori* IgG may accurately stratify those at higher risk of gastric cancer.[28] Recently, trefoil factor 3, was shown to be a more sensitive candidate than PG1 for detection of gastric cancer, but together these biomarkers have a synergistic effect.[29, 30]

Most recently, potential volatile biomarkers within the exhaled breath of patients with oesophagogastric cancers have been identified. Kumar *et al* identified 12 VOCs (pentanoic acid, hexanoic acid, phenol, methyl phenol, ethyl phenol, butanal, pentanal, hexanal, heptanal, octanal, nonanal, decanal) present in significantly higher concentrations in oesophagogastric cancer patients (n=81) than healthy controls (n=129).[31] The area under the receiver operating characteristic curve (ROC) was 0.97 for oesophageal cancer and 0.98 for gastric cancer. A model and validation subset were conducted with 8 significant predictors (decanal, nonanal, phenol, ethyl phenol, methyl phenol, hexanoic acid, heptanal, butyric acid) demonstrating an area under the ROC curve of 0.92 and 0.87 respectively. The model achieved a diagnostic accuracy of 86% and the validation subset 83.3%.[31] A subsequent multicentre validation study with 335 patients (oesophagogastric cancer n=163, controls n=172) established a diagnostic exhaled breath model based on five distinct volatile organic compounds (butyric acid, pentanoic acid, hexanoic acid, butanal and decanal). This study demonstrated a sensitivity of 80%, specificity of 81% and area under the ROC curve of 0.85.[32] Volatile short chain fatty acids and aldehydes have proven to be important potential breath biomarkers. Supporting studies exploring the direct association between VOCs and oesophagogastric tumours have shown elevated levels of acetic-, butyric-, pentanoic- and hexanoic acids in the *ex vivo* tumour headspace.[33] *In vivo* VOC analysis of endoluminal tumour headspace showed increased butyric acid ($p<0.001$) and pentanoic acid ($p=0.005$) in oesophagogastric cancer patients (n=25) compared to healthy controls (n=20), with an area under the ROC of 0.80.[33] However, breath analysis is not without challenges. Breath is a complex biological matrix which encompasses information about patients' systemic metabolic status. The removal of 'waste' metabolites via the bloodstream, transfer across the alveoli surface and exhalation provides the opportunity for them to be detected non-invasively within breath. However, this introduces many potential confounding factors from the patient (concurrent diseases, diet, medications) and environment (ambient contaminants within inhaled breath). Refining the metabolite profile, and identifying the source is key in demonstrating potential biomarkers specific for disease types. Another challenge is the low concentrations of target metabolites requiring specific handling and processing of the samples, use of costly mass spectrometry platforms and expertise in interpretation of the end data.

Notwithstanding, this work collectively supports breath analysis as a non-invasive novel platform for the discovery of candidate diagnostic biomarkers for oesophagogastric cancer. The ability to augment the production of cancer specific VOCs, may however offer an attractive method of improving the diagnostic performance of the proposed breath test, whilst at the same time mitigating unwanted sources of

variability. Bacteria within the upper gastrointestinal tract could represent the origin of VOC as well an accessible target for their augmented production.

1.3 MECHANISMS OF VOC AUGMENTATION: OESOPHAGOGASOGASTRIC CANCER MICROBIOME AND CANCER CELLS

1.3.1 Oesophagogastric Cancer Microbiome

The microbiome encompasses the genetic material of bacteria, viruses, fungi and protozoa harbouring the human body, which accounts for more than 130 times the quantity of the human genome.[34] Equally, the inter-individual variability of the microbiome is up to 90% compared to the 99.9% similarity in the human genome.[34] This highlights the potentially significant contribution of the microbiome to disease states. There is an increasing interest in the gut microbiome, its relationship with the oral cavity and the potential role as a biomarker in disease.

Over 100 trillion bacteria populate the human body, with some of the most densely populated bacteria within the oral cavity (700 species) and specific microbiota associated with specific organ function.[35] The gut microbiome is specifically developed in an environment enriched for nutrient and drug metabolism, maintaining adequate immune function and a functional mucosal barrier. Studies over the last few decades have suggested the role of the gut microbiome in the development of benign and malignant gastrointestinal disease. The microbiome within the body may be directly influenced by disease, homeostatic changes, medications such as immunosuppressants and antibiotics, diet and stress.[36, 37] The communication between the oral cavity, oesophagus and stomach allows translocation of bacteria in two ways; following the natural anatomical path via swallowing, and pathological path with reflux altering the microenvironment.[38] There is new evidence suggesting bacteria may be the drivers of cancer formation, by directly promoting DNA damage, or indirectly by localised inflammation.[39] The upper gastrointestinal microbiome is not as well characterised compared to the microbiome of the colon and rectum.[40, 41] The stomach has traditionally been considered a 'sterile' environment due to its ability to maintain a hyper-acidic state where many bacteria are unable to survive. Selective bacteria flourishing in this environment are considered fastidious, meaning they have unique properties to endure these environmental conditions.

The microbial ecosystem is shared between the oesophagus and stomach. The core microbial composition comprises five predominant phyla: the most abundant phylum are *Firmicutes* (Gram-positive; 83%) with *Streptococcus viridans* most commonly residing in both the normal oesophagus and oropharynx. Other phyla include *Proteobacteria* (8%), *Bacteroidetes* (7%), *Actinobacteria* (2%), and *Fusobacteria* (1%) which are predominantly Gram-negative.[35, 42-45] Studies have demonstrated progressive microbial changes from benign to premalignant to cancer states: increase in *Lactobacillus* from gastrooesophageal reflux disease to oesophageal adenocarcinoma, and a change in microbial profile from superficial gastritis progressing to gastric cancer.[42, 46] Furthermore, a recent increasing interest in the link between the oral and the tumour microbiome presents novel non-invasive opportunities for diagnostic biomarkers. Microbes associated with oral pathology such as *Tannerella forsythia* have been linked with an increased risk of developing oesophageal adenocarcinoma.[47]

Other studies have characterised a panel of 13 bacteria enriched in gastric cancer, related to the commensal organisms inhabited in the oral cavity.[48] Most recently, Sun *et al* demonstrated a sensitivity of 97% in screening gastric cancer using the oral microbiome.[38] However, the association between the microbiome and the development of oesophagogastric cancer remains incompletely understood.

Microbial assessment of oesophagogastric cancer in our own patient cohort supported literature findings. Predominant bacteria associated with cancer were found through 16S rRNA gene sequencing: *Firmicutes* (*Lactobacillaceae*, *Streptococcaceae*, *Staphylococcaceae*, *Eubacteriaceae* families), *Proteobacteria* (*Escherichia coli*, *Klebsiella pneumonia*, *Pasteurellaceae*) and *Fusobacteria* (*Fusobacteriaceae*, *Leptotrichiaceae*) (Published Thesis, Mina Adam; Imperial College London 2019). Direct manipulation of the microbiome and its potential to alter the molecular mechanisms promoting carcinogenesis may have useful implications for future clinical practice providing a platform for novel diagnostic and treatment strategies.

1.3.2 Bacterial augmentation of Volatile Organic Compounds

Exhaled breath tests are well established in clinical practice with the urease breath test diagnostic of *Helicobacter pylori* and the nitrogen breath test for asthma.[49] These end metabolites are produced from intrinsic metabolic processes. More recently, intrinsic metabolic pathways have been exploited to establish augmented exhaled breath tests. The diagnosis of small intestinal bacterial overgrowth relies on proximal gut bacterial metabolism of consumed glucose to produce hydrogen detected within the breath. Based on this principle, a pilot study by our group demonstrated that simple dietary nutrients can be used to increase target VOC production. *In vitro* micro-aerobic studies with cancer-related bacteria, *Escherichia coli* (NCIMB 9552), demonstrated an augmentation of acetic acid following introduction of a nutrient stimulant mix. With refinement of the nutrient stimulant mix, further high throughput biotransformation of *Escherichia coli* (NCIMB9552), *Lactobacillus fermentum* (NCIMB11840), *Streptococcus salivarius* (NCIMB701779), *Klebsiella pneumonia* (NCIMB13281) and *Streptococcus anginosus* (NCIMB702496) showed elevated target VOC levels compared to controls (Published Thesis, Mina Adam; Imperial College London 2019). A further clinical study introducing the nutrient stimulant drink to patients with oesophagogastric cancer, demonstrated a 1.6-fold increase in butyric acid after 30 minutes (Published Thesis, Mina Adam; Imperial College London 2019).

These initial *in vitro* and *in vivo* studies support the basis of an augmented breath test exploiting the tumour-associated microbiome in oesophagogastric cancer.

1.4 THESIS HYPOTHESIS AND AIMS

Hypothesis

Stimulation of the tumour-associated microbiome will result in augmented VOC production that may enhance the detection of oesophagogastric cancer.

Aims

The primary aim of this thesis is to investigate the role of the oral and tumour-associated microbiome in the augmented production of VOCs in oesophagogastric cancer.

Secondary aims were to:

1. Systematically review existing literature relating to: (i) the oesophagogastric cancer associated microbiome, and (ii) augmentation of short chain fatty acids by bacterial metabolism.
2. Describe the oral and tumour-associated microbiome of oesophagogastric cancer.
3. Measure VOC production from microbial cultures derived from oral and oesophagogastric tumour samples.
4. Characterise augmented VOC production from microbial cultures in response to a nutrient stimulant challenge.
5. Define the augmented exhaled VOC response to an oral stimulant challenge in oesophagogastric cancer patients.

2. CHAPTER 2: EVALUATION OF THE OESOPHAGOGASTRIC CANCER ASSOCIATED MICROBIOME: A SYSTEMATIC REVIEW AND QUALITY ASSESSMENT

2.1 INTRODUCTION

The gastrointestinal microbiome and its role in cancer pathogenesis may offer an opportunity for new advancements in diagnostics, therapeutics, and prognostication. *Helicobacter pylori*'s established role in the pathogenesis of gastritis, peptic ulcer disease and transformation to cancer has paved the way for exploring the wider influence of bacteria on cancer.[50] Previous studies have identified characteristic bacterial phenotypes related to intestinal cancers, including oesophageal, gastric and colorectal cancer. Concurrent studies also found that the oral microbiome of gastric, colorectal, pancreatic and lung cancer patients was unique, offering a potentially non-invasive diagnostic window into the digestive tract.[38, 51-53] In addition, pre-operative saliva samples have been predictive of anastomotic leak and pneumonia following oesophagectomy.[54, 55] Advances in sequencing techniques have led to a better understanding of the diversity and composition of microbial communities, but it remains unclear whether this change is the cause or effect of a disease process.

2.2 AIM

This systematic review aims to:

- (i) Evaluate the current published literature relating to the microbiome in oesophagogastric cancer.
- (ii) Identify bacteria associated with oesophageal and gastric cancer.
- (iii) Establish the influence of sample location and analytical method.

2.3 METHODS

Search strategy

A systematic search was performed using MEDLINE (OvidSP), EMBASE (OvidSP) and Cochrane databases to identify potentially relevant articles published between 1966 to 21st December 2020. A search strategy formed with keywords and MeSH headings relating to “oesophageal cancer,” “gastric cancer” and “microbiome” used in combination with the Boolean operators AND and OR. Details of the full search strategy used is provided in Appendix 13.1. Other databases including the World Health Organisation International Clinical Trials Registry, ClinicalTrials.gov, ISRCTN Register, and PROSPERO were reviewed to identify ongoing studies.

Eligibility assessment and data extraction

Inclusion criteria included: studies reporting bacteria identified in patients with oesophageal and gastric cancer (any histological type) and compared to an appropriate control group of healthy subjects or patients with benign conditions of the upper gastrointestinal tract. All bio-sample types and methods for bacterial identification were included. Review articles, case reports, editorials, opinions, conference abstracts and articles not written in the English language were excluded. Two independent reviewers (BV, MT) screened all titles and abstracts to identify relevant articles for full text review. A third reviewer (PB) resolved any disagreements in article selection. Reference list of all selected articles were also searched to identify additional relevant articles. Data extraction from full text review included: study design, cohort demographics, specimen type, microbiome assessment method, diversity characteristics (including α - and β diversity metrics), bacteria identified at all taxonomic levels with P-values <0.05.

Definitions of groups

Subjects were classified into three groups: (i) patients with histologically proven oesophageal and gastric cancer; (ii) benign disease of the upper gastrointestinal tract, including superficial gastritis, atrophic gastritis, and (iii) healthy controls with no known pathology of the upper gastrointestinal tract. Premalignant conditions such as Barrett's oesophagus or intestinal metaplasia were not included in the analysis.

Outcomes

The primary outcome was to identify bacterial enrichment specific to oesophagogastric cancer in comparison to a relevant control group. All bacteria identified were recorded, focusing on those reported in five or more studies. Other microorganisms such as fungi, viruses and protozoa were not recorded. Secondary outcomes included: (i) investigating the relationship between the oral and cancer tissue microbiome; (ii) assessing the diagnostic performance of the cancer-associated microbiome; (iii) describing the microbial diversity, and (iv) methodology appraisal.

Quality assessment

The Newcastle-Ottawa Scale was used to assess the quality and risk of bias for case-control and cohort studies. Two independent researchers (MT, BV) graded the studies across three domains: selection of the study groups (adequacy of case definition, representativeness of the cases, selection of controls, definition of controls), comparability of the groups (based on design and accountability of confounding factors), and outcome of the studies (exposure, assessment methods, follow-up).

Studies were also evaluated using the Strengthening The Organisation and Reporting of Microbiome Studies (STORMS) checklist, recently developed for reporting on the quality of human microbiome studies [56]. It comprises of a comprehensive structured 17-item checklist intended to standardise reporting of important quality and methodology parameters.

2.4 RESULTS

Study selection

A total of 5861 articles were identified. After removing duplicates, 4747 articles underwent title and abstract screening. Fifty-six studies were selected for full text review, of which 10 were excluded as they were either a single arm study (without control group) or because they did not report relevant information (e.g., named bacteria). Final analysis included 46 studies: 31 studies assessing gastric adenocarcinoma [38, 44, 46, 48, 57-83], and 17 studies assessing oesophageal carcinoma (adenocarcinoma and/or squamous cell carcinoma)[47, 55, 69, 72, 84-96] (Figure 1). Two studies reported data from both oesophageal and gastric cancers.[69, 72]

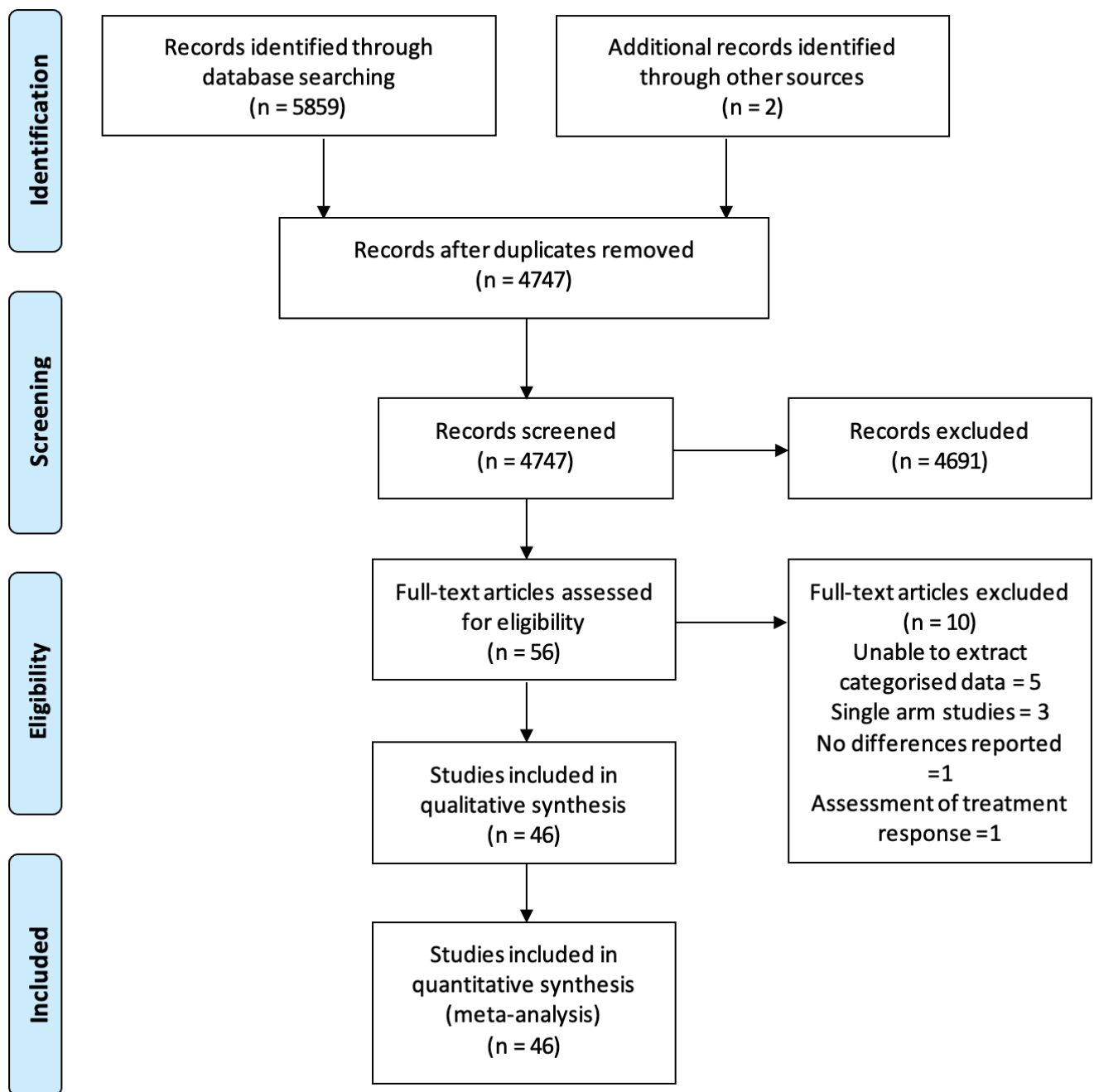


Figure 1. Flowchart describing article selection for the assessment of the oesophagogastric cancer microbiome.

Gastric adenocarcinoma

Study characteristics: Thirty-one studies included analysis of 3381 patients with either: gastric adenocarcinoma (n=1828); benign disease (n=654), or a healthy upper gastrointestinal tract (n=899). Full details of study characteristics are presented in Table 2. The majority of studies were conducted in Asian populations (n=26). With the exception of one study (retrospective review of registered database)[58], all studies were prospective case-control studies, where the control group was formed of either unrelated healthy (non-cancer) patients or healthy tissues samples (adjacent to the tumour) acquired from cancer patients.

Sample type: Microbial assessment was performed in various biological matrices, including upper gastrointestinal tract tissue (n=21); oral samples (saliva, n=3; tongue coating, n=3; subgingival plaque, n=11); gastric/oesophageal fluid (n=2); faeces (n=2), and serum (n=1).

Method for bacterial identification: Significant heterogeneity was observed in the sequencing technique used to examine the microbiome. The majority of studies (n=27) used 16S rRNA gene sequencing predominantly targeting the V3-4 hypervariable region. Five studies that utilised 16S rRNA gene sequencing did not specify which variable region was targeted.[58, 59, 65, 71, 81] Two studies employed primary bacterial culture from acquired tissue samples with subsequent 16S rRNA gene V3-4 sequencing [55, 57] and two studies utilised shotgun sequencing and terminal restriction fragment length polymorphism.[44, 48]

Bacteria enriched in Gastric adenocarcinoma: Taxonomic classifications to the genus level were reported in the majority of studies. A total of 239 bacteria were enriched in gastric cancer at the following levels: 12 phyla, 8 classes, 16 orders, 33 families, 95 genera, 73 species and 2 strains. Bacteria that were reported by three or more studies (36/239, including all taxonomic levels) are shown in Figure 2.

All six genera observed to be enriched in gastric adenocarcinoma belonged to the three predominant phyla. *Lactobacillus* (Phyla: *Firmicutes*) was reported in 14 studies (45%) in multiple biological matrices: tumour tissue (n=11); tongue coating (n=1); faeces (n=1), and gastric/oesophageal fluid (n=1). *Streptococcus* (Phyla: *Firmicutes*) was identified in 10 studies: tumour tissue (n=7); tongue coating (n=1); faeces (n=1), and gastric/oesophageal fluid (n=1). *Prevotella* (Phyla: *Bacteroidetes*) was present in seven studies: tumour tissue (n=5); saliva (n=5), and faeces (n=1). *Fusobacterium* (Phyla: *Fusobacteria*) was reported in tumour tissue (n=6) and from gastric/oesophageal fluid (n=1). *Veillonella* (Phyla: *Firmicutes*) was found in seven studies and was reported in various biological matrices including tumour tissue (n=3), faeces (n=2), subgingival plaque (n=1) and gastric/oesophageal fluid (n=1). *Bacteroides* (Phyla: *Bacteroidetes*) was also found in four studies: tumour tissue (n=2); in gastric/oesophageal fluid (n=1), and; in serum (n=1).

Bacteria enriched in benign gastric conditions: The benign group demonstrated an enrichment of 156 bacteria reported at the following levels: 7 phyla, 5 classes, 14 orders, 22 families, 54 genera, 50

species and 4 strains. Four bacteria were reported in three or more studies. No bacteria were reported in five or more studies.

Bacteria enriched in healthy control group: The healthy control group showed enrichment of 95 bacteria reported at the following levels: 13 phyla, 2 classes, 3 orders, 5 families, 50 genera, and 22 species. Five studies demonstrated a higher level of the *Neisseria* genus in tongue coating samples (n=3); tissue (n=1), and gastric/oesophageal fluid (n=1).

Microbial diversity metrics: 28 studies (90%) assessed microbial diversity. The majority of studies employed more than one metric. Alpha diversity, a measure of the number (richness) of species within a sample, was reported in 27 studies. Richness indices reported included: Shannon index (n=23); OTUs (n=14); Chao1 (n=14); Simpson Index (n=8); Phylogenetic Diversity tree (n=8), and ACE (n=7). All alpha diversity metrics reported different outcomes with either an increase, decrease or no change in the species richness. Overall, six studies demonstrated an increase in alpha diversity, six studies reported a decrease, five studies showed no difference and ten studies reported at least two changes. Beta diversity, a measure of the difference in composition, was measured in 23 studies; principal coordinate analysis was used in 20 studies, weighted and unweighted UniFrac distance was employed in 13 studies and Bray Curtis in 12 studies. Fifteen studies reported a significant difference between cancer and benign/control samples. Six studies showed no significant difference, and one study reported conflicting outcomes. Evaluation of diversity metrics has shown conflicting results for alpha diversity but trends towards a significant difference in beta diversity between cancer and non-cancer samples.

Table 2. Study characteristics of articles reporting on the microbiome of gastric adenocarcinoma compared to a non-cancer group.

Author	Year	Country	Specimen type	Gastric adenocarcinoma			Control group	Control			Microbiome assessment method	MCC	No. of genera enriched (C/B/Hc)	NAT	PPI	Abx	Ref.
				No. of patients	Age (yrs)	Gender (male)		No. of patients	Age (yrs)	Gender (male)							
Sjostedt	1985	Sweden	Saliva, OG fluid	10	-	-	B, HC	30, 30	-	-	Culture	-	8/9/9	-	No	No	[57]
Dicksved	2009	Sweden	Tissue	10	71	8	HC	5	73	3	T-RFLP	-	-	-	-	No	[44]
Seo	2014	Korea	Tissue	16	60	12	AHT	-	-	-	RNA seq database	-	-	-	-	-	[58]
Aviles-Jimenez	2014	Mexico City	Tissue	5	71	1	B	5	44	0	16S gene, microarray hybridization	Yes	0/4/-	-	No	No	[59]
Eun	2014	Korea	Tissue	11	66	6	B	10	50	4	16S gene V5, pyroseq	-	2/0/-	-	No	No	[60]
Hu	2015	China	Tongue coating	74	57	37	HC	72	55	35	16S gene V2-V4, Illumina seq	-	0/-/4	No	-	No	[61]
Wang	2016	China	Tissue	103	-	-	B	212	-	-	16S gene V1-V3, pyrosequencing	-	3/0/-	-	No	No	[62]
Yu	2017	China/Mexico	Tissue	160	63	109	AHT	-	-	-	16S gene V3-V4, MiSeq	-	5/-/0	No	-	-	[63]
Li	2017	Hong Kong	Tissue	7	53	5	B, HC	9, 8	50, 49	2, 3	16S gene V3-V4, Solexa Illumina seq	-	9/2/8	-	No	No	[64]
Castano-Rodriguez	2017	Malaysia	Tissue	12	62	4	B	20	50	11	16S gene, MiSeq	-	24/8/-	No	No	No	[65]
Ferreira	2018	Portugal	Tissue	54	59	32	B	81	44	79	16S gene V5-V6, NGS	Yes	6/5/-	-	No	No	[66]
Sun	2018	China	Saliva, SP	37	-	-	HC	-	-	-	16S gene, MiSeq	Yes	5/-/7	-	-	No	[38]
Hu	2018	China	Gastric fluid	6	61	5	B	5	55	4	Shotgun, HiSeq	Yes	4/4/-	No	No	No	[48]
Wu	2018	China	Tongue coating	57	59	40	HC	80	55	50	16S gene V4, pyrosequencing	Yes	6/-/23	No	No	No	[67]
Hsieh	2018	Taiwan	Tissue	11	69	5	B	9	32	3	16S gene V3-V4, MiSeq	-	3/0/-	-	-	-	[68]
Coker	2018	China	Tissue	20	-	-	B	44	-	-	16S gene V4, N-W algorithm	Yes	6/2/-	No	No	No	[46]
Shao	2019	China	Tissue	34	62	-	AHT	-	-	-	16S gene V4 + miniseq	-	17/-/0	-	-	-	[69]
Gunathilake	2019	South Korea	Tissue	268	54	172	HC	288	52	181	16S gene V3-V4 + MiSeq	Yes	2/-/2	-	-	-	[70]
Liang	2019	China	Faeces	20	52	-	HC	22	53	-	16S gene +MiSeq	-	3/-/1	No	No	No	[71]
Kageyama	2019	Japan	Saliva	10	71	6	HC	118	66	84	16S gene V1-V2 + Ion PGM Hi-Q Seq	-	-	No	-	No	[72]
Chen	2019	China	Tissue	62	60	46	AHT	-	-	-	16S gene V4-V5 HiSeq	Yes	9/-/0	No	No	No	[73]
Dong	2019	China	Serum	71	59	60	HC	13	-	-	16S gene V1-V2 HiSeq	Yes	3/-/2	-	No	No	[74]
Liu	2019	China	Tissue	276	61	195	AHT	-	-	-	16S gene V3-V4 MiSeq	-	7/-/0	No	No	No	[75]
Qi	2019	China	Faeces	116	58	96	HC	88	46	53	16S gene V3-V4 MiSeq	-	11/-/6	No	-	No	[76]
Wang	2020	China	Tissue	29	57	18	B, HC	21, 30	45, 46	9, 15	16S gene V4 MiSeq	Yes	5/4/2	No	No	No	[77]
Wang	2020	China	Tissue	60	60	44	B	60	52	41	16S gene V3-V4 HiSeq	-	16/20/-	-	No	No	[78]
Spiegelhauer	2020	Denmark	Tissue	12	62	7	B	22	48	4	Culture + 16S gene V3-V4 HiSeq	Yes	1/0/-	No	No	No	[79]
Gantuya	2020	Mongolia	Tissue	48	-	-	B, HC	60, 20	-	-	16S gene V3-V4 + MiSeq	-	8/0/0	-	No	No	[80]

Wu	2020	China	Tissue	18	63	15	B	32	62	24	16S gene + HiSeq	-	11/16/-	No	No	No	[81]
Xu	2020	China	Tongue coating	181	64	146	HC	112	64	91	16S gene V3-V4 + MiSeq	Yes	12/-/14	No	No	No	[82]
Dang	2020	China	Tissue	30	-	-	B	17	-	-	16S gene V3-V4 + MiSeq	-	4/0/-	-	-	-	[83]

OG, oesophagogastric. SP, Subgingival plaque. C, cancer. B, benign, HC, healthy control. AHT, adjacent healthy tissue. 16S, 16S rRNA gene. N-W, Needleman-Wunsch. MCC, multiple comparisons correction. NAT, neoadjuvant therapy. PPI, proton pump inhibitor therapy. Abx, Antibiotic therapy. Y, yes. N, no.

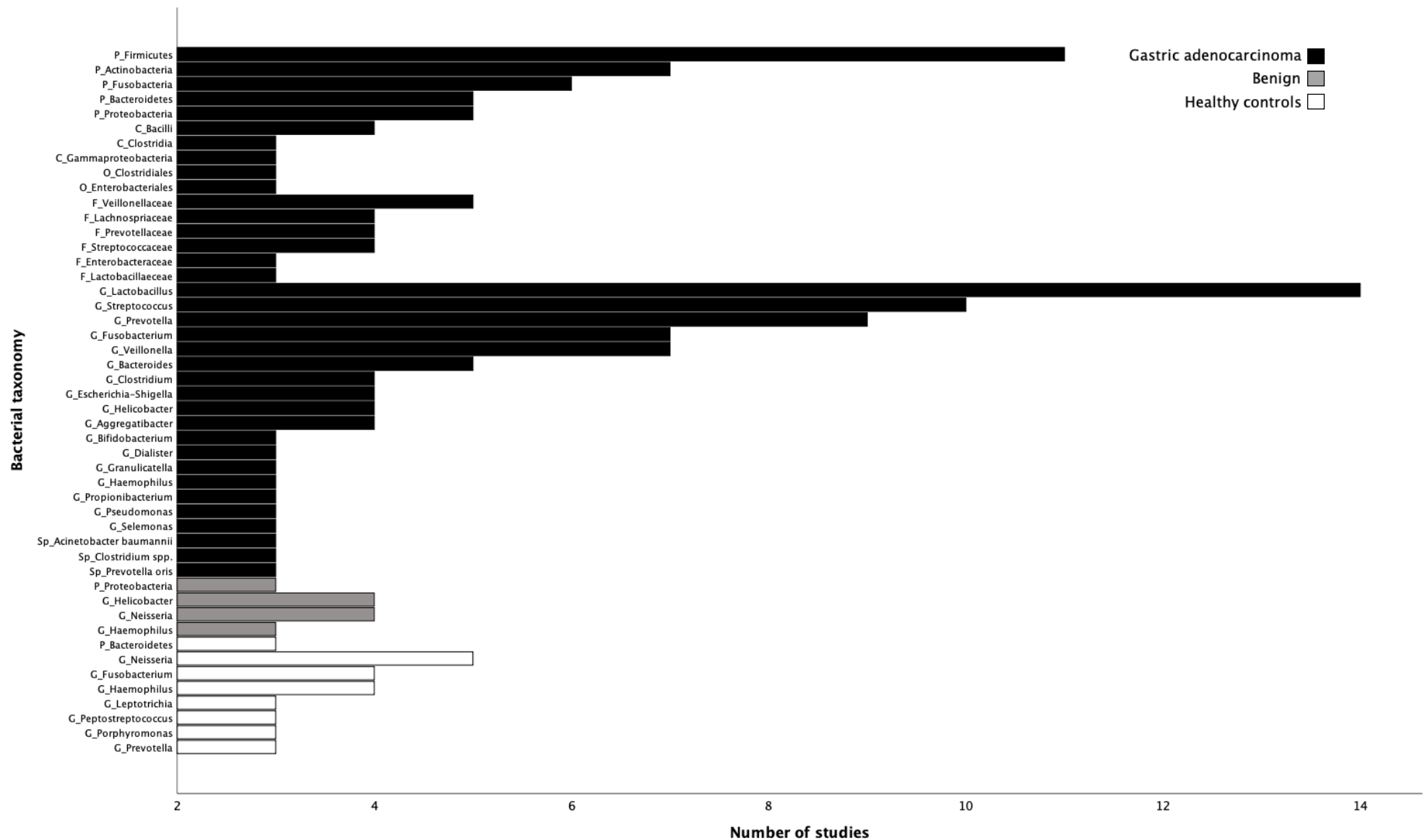


Figure 2. Bacteria enriched in gastric adenocarcinoma, benign disease and healthy controls. Bacteria identified to be significant in 3 or more studies out of 31 studies were included. P = phyla, C = class, O = order, F= family, G = genus, Sp = species. Five genera, *Lactobacillus*, *Streptococcus*, *Prevotella*, *Fusobacterium* and *Veillonella* were demonstrated in higher abundance in gastric adenocarcinoma.

Oesophageal carcinoma

Study characteristics: Seventeen studies included the analysis of the oesophageal cancer microbiome, including oesophageal squamous cell carcinoma (SCC, n=7); oesophageal adenocarcinomas (OAC, n=4) and mixed SCC/OAC samples (n=6) (Table 3). A total of 2735 patients were included in the analysis: 1517 patients with SCC (55%), 514 OAC (19%); 656 healthy controls (24%); and 48 with benign disease (2%). Twelve of the 17 studies originated from Asian populations. All studies were prospective case-control studies.

Sample type: Primary sites for sample collection were from the upper gastrointestinal tract and the oral cavity. Tissue biopsy (n=9) and saliva (n=9) including mouthwash samples were the most common, with a single study sampling subgingival plaque.

Method for bacterial identification: Thirteen of the 17 studies identified bacteria by 16S rRNA gene sequencing with different targeted variable regions, V3-4 being the most common (n=7). Other methods employed were q-PCR, RT-PCR, and bacteria culture methods.

Bacteria enriched in oesophageal carcinoma: No common trend was observed at any taxonomic level for studies that evaluated bacteria related to oesophageal carcinomas. A total of 80 oesophageal SCC specific taxa were reported: 4 phyla, 4 classes, 8 orders, 17 families, 41 genera, and 6 species. Limited OAC studies showed the enrichment of 35 bacteria: 3 classes, 2 orders, 2 families, 16 genera and 12 species. The comparison control group demonstrated a higher diversity of 128 bacteria: 8 phyla, 7 classes, 12 orders, 26 families, 51 genera and 24 species. Of note, oesophageal SCC reported on two genera which were identified in three studies. *Streptococcus* was enriched in SCC tissue, saliva and oral mucosal swabs in three studies [84, 90, 92], whilst *Prevotella* was enriched in saliva from two studies and tissue in one study.[84, 90, 93] There were no significant findings for OAC or benign disease groups. A single genus, *Neisseria* was reported in healthy controls in three studies assessing saliva.[47, 84, 93] With significant methodological heterogeneity, robust conclusions could be drawn from these studies.

Table 3. Study characteristics of articles reporting on the microbiome of oesophageal cancer compared to a non-cancer group.

Author	Year	Country	Specimen type	Histological subtype	Oesophageal carcinoma			Control group	Control			Microbiome assessment method	MCC	No. of genera enriched (C/B/Hc)	NAT	PPI	Abx	Ref.
					No. of patients	Age (yrs)	Gender (male)		No. of patients	Age (yrs)	Gender (male)							
Chen	2015	China	Saliva	SCC	87	65	59	HC	85	66	62	16S gene V3-V4, pyrosequencing	-	7/-/26	No	-	No	[84]
Nasrollahzadeh	2015	Iran	Tissue	SCC	37	65	18	B, HC	17, 37	64, 62	16, 18	16S gene V3-V4, GS-FLX	Yes	0/0/0	-	-	-	[85]
Yamamura	2016	Japan	Tissue	SCC, Adeno	12, 300	66	287	HC	-	-	-	qPCR	-	0,0/7	No	-	No	[86]
Elliott	2017	UK	Tissue	Adeno	19	70	15	HC	20	57	7	16S gene V1-V2, MiSeq	Yes	0/-/0	Mixed	-	-	[87]
Peters	2017	USA	Mouthwash	Adeno	25, 81	67, 68	10, 75	HC	210	68	168	16S gene V4, MiSeq	-	1/-/4	-	-	-	[47]
Shao	2019	China	Tissue	SCC	61	62	-	-				16S gene V4, miniSeq	-	2/0	-	-	-	[69]
Kageyama	2019	Japan	Saliva	Unspecified	12	68	8	HC	118	66	84	16S gene V1-V2, Ion PGM Hi-Q Seq	-	0/-/0	No	-	No	[72]
Wang	2019	China	Saliva	SCC	20	66	14	HC	21	65	12	16S gene V3-V4 MiSeq	-	3/-/2	No	-	No	[88]
Yamamura	2019	Japan	Tissue	SCC	551	66	459	-	-	-	-	qPCR	-	0/0	Mixed	-	-	[89]
Xu	2020	China	OM swab	SCC	31	-	24	-	-	-	-	16S gene V3-V4, Ion S5 TM XL	-	3/0	Yes	-	No	[90]
Yuda	2020	Japan	Saliva	SCC, Adeno	567, 78	64	572	-	-	-	-	Culture	-	0/0	Mixed	-	No	[55]
Peter	2020	USA	Tissue	Adeno	10	65	9	HC	10	52	6	16S gene V4, Miseq	Yes	0/-/3	-	Mixed	No	[91]
Li	2020	China	Tissue	SCC, Adeno	17, 11	61, 61	12, 8	HC	16	59	10	16S gene V3-V4, Miseq	-	11,8/8	No	-	No	[92]
Zhao	2020	China	Saliva	SCC	39	60	23	HC	51	49	23	16S gene V3-V4, Miseq	Yes	24/-/11	No	-	No	[93]
Zhou	2020	Australia	Tissue	Adeno	6	67	6	B, HC	31, 16	-, 52	-, 2	16S gene V1-V3, Miseq	-	3/4/3	No	-	No	[94]
Lopetsu	2020	Italy	Tissue	Adeno	6	56	4	HC	10	52	6	16S gene V3-V4, Miseq	-	5/-/3	No	N	No	[95]
Kawasaki	2020	Japan	SP	SCC, Adeno	58, 3	66	47	HC	62	70	50	RT-PCR	-	0,0/0	No	-	No	[96]

OM, oral mucosa. SP, Subgingival plaque. SCC, squamous cell carcinoma. Adeno, adenocarcinoma. C, cancer. B, benign, HC, healthy control. 16S, 16S rRNA gene. MCC, multiple comparisons correction. NAT, neoadjuvant therapy. PPI, proton pump inhibitor therapy. Abx, Antibiotic therapy. Y, yes. N, no

Evaluation of diagnostic performance

The diagnostic performance of the microbiome was assessed by Sun *et al.*, who reported a 97% sensitivity of saliva and subgingival plaque for gastric cancer screening.[38] No other clinically relevant applications of the microbiome were determined from studies presented herein.

Influence of sample origin

There was insufficient data to compare the relationship between the tumour tissue and oral microbiome of patients with oesophagogastric. No studies assessed these two sample locations in the same cohort of patients.

Quality of assessment

The Newcastle Ottawa Scale for case-control and cohort studies assessed the quality of reporting.[97] Twenty-one (46%) studies scored the maximum of 9 stars; two studies (4%) scored 8 stars, 22 (48%) scored 7 stars and one study scored 6 stars (2%). All studies reported patient selection representative of the community population, apart from one study which enrolled hospitalized patients. Microbiome results were reported accurately in accordance with the outcome measures. Although age and gender were provided for all groups, comparability between groups, including statistical analysis of baseline characteristics were poorly reported (Appendix 13.2 and 13.3).

The STORMS checklist demonstrated clear reporting of the rationale of the study (100%), detailed sample collection (100%), storage (85%), and 16S rRNA gene sequencing strategies (96%). Few studies used positive controls (mock communities, 4%), discussed batch effects (11%), and contamination within DNA extraction kits/reagents which are prevalent in low biomass samples (30%). Important microbial influencing factors such as antibiotic use (80%) and matching criteria (59%) were not reported in all studies. Although it is strongly encouraged to deposit sequence reads in an established publicly available repository such as NCBI, this was done in only 41% of studies.

Full details of bacteria identified, and quality assessment scoring can be provided on request: Supplementary files.

2.5 DISCUSSION

The principal findings were: (i) the enrichment of select bacteria in gastric adenocarcinoma compared to benign disease and healthy controls, with distinctive bacterial profiles observed at all taxonomic levels; (ii) enrichment of the genera *Lactobacillus* and *Streptococcus* in gastric adenocarcinoma; (iii) higher abundances of *Firmicutes* in various bio-samples including tumour tissue, saliva and gastric fluid; (iv) similarities observed in bacterial taxa found in different bio-samples may support non-invasive approaches for characterising the microbiome, and; (v) significant methodological heterogeneity in standards of reporting.

Enrichment of the genera *Lactobacillus* and *Streptococcus* were common findings in 11 studies (35%) evaluating gastric adenocarcinoma. Although this review collectively supports an important alteration in species composition, there is disagreement between species diversity. Most studies used more than one alpha diversity metric producing conflicting results. Although it is becoming increasingly evident that microbial alteration in disease is important, assessment of the functional capabilities provides more valuable information. The functional ecology describes the active contribution of microbes to a system. This doesn't necessarily reflect the species composition. Species may be present; however, they can be functionally redundant. Therefore, metagenomic studies are recommended as pathway analysis may be helpful in targeting treatments by better understanding the cancer-microbe-immune axis. One study by Hu *et al.*, presented significantly different bacteria-enriched pathways in gastric cancer compared to superficial gastritis. Cancer-specific metabolites were generated from pathways related to amino acid metabolism (L-arginine, L-ornithine) and the pentose phosphate pathways.[48]

The role of the microbiome in the pathogenesis of gastric cancer was demonstrated by Wang *et al.*, who identified characteristic bacterial variation that was related to disease progression from chronic gastritis to intestinal metaplasia to gastric cancer.[77] Specifically, *Actinobacteria*, *Bacteroides*, *Firmicutes* and *Fusobacteria* were enriched in premalignant and cancer states with candidate OTUs demonstrating an area under the curve >0.9 for gastric cancer detection.[77] *Firmicutes* were also reported in higher abundances in different biological matrices including the tissue, gastric fluid and saliva of gastric cancer patients. Similarities in bacterial taxa found in different biological samples may support future non-invasive methods for characterising in oesophagogastric cancer. Sun *et al.*, reported that saliva and subgingival plaque microbiome could detect gastric cancer with 97% sensitivity.[38]

A limited number of studies with variable methodologies meant that fewer trends were observed in the oesophageal cancer microbiome. From 17 studies, only three studies identified the same two genera in oesophageal SCC: *Streptococcus* and *Prevotella*. [84, 90, 92, 93] No bacterial changes were consistently observed in more than three studies. *Fusobacterium nucleatum*, which predominantly inhabits the oral cavity and has been linked to tumour aggression and shorter survival from oesophageal SCC.[86]

Forty-one of the 46 studies identified by this review were published in the last five years highlighting how interest in the upper gastrointestinal cancer microbiome is only a relatively recent occurrence. During the emergence of this field there has been significant heterogeneity between the methodologies used by individual studies. Whilst 16S rRNA gene sequencing was the most common method for defining the microbiome used in studies identified by this review, there were differences in the variable region targeted. As different variable regions amplify different bacteria, there is a risk of taxonomic bias. Studies that also included culture methods would introduce further selection bias. It is therefore recommended that a standardised methodology for sample acquisition and microbiome assessment

including DNA extraction and sequencing, should be considered. The diversity of methods used may explain the observed variation of results.

Existing checklists for the assessment of study reporting and quality do not encompass specific aspects of study design are important to human microbiome research. Many checklists, including the Newcastle Ottawa Scale that was used to assess studies included in this review, fail to assess both patient specific (e.g., diet, co-morbidity, medications) and analytical (e.g., sample handling, analysis, and data processing) factors that have the potential to influence findings. Recently a consortium of researchers have proposed a checklist to improve reporting of human microbiome research.[56] Whilst the STORMS checklist has not yet been validated, it can serve as guide to researchers within the field. The checklist comprises key features important in reporting high quality and reproducible microbiome research. For studies included in this review it highlighted: (i) deficiencies in eliminating sources of contamination by using controls and assessing DNA extraction kits; (ii) the need to control for patient factors influencing the microbiome, and (iii) heterogeneity in laboratory methods used including target variable regions. Whilst it is beyond the scope of the current to establish similar guidance, important considerations for the specific analysis of the upper gastrointestinal microbiome are presented in Table 4 and are intended to serve as a guide to future investigators.

Limitations of this study include a wide methodological variation making it impossible to draw robust conclusions from the combined outcomes of selected studies. Over three quarters of included studies recruited less than eighty cancer patients suggesting that they may have been underpowered to draw definitive conclusions. No study presented a formal power calculation. Furthermore, only 15 studies out of 46 (33%) used multiple comparison correction. Studies were predominantly from Far Eastern centres and as such it is uncertain how translatable results may be for patient from other global regions. Microbial profiles are influenced by many environmental factors, and geographical location encompasses many of these. With the recent increase in the incidence of oesophageal adenocarcinoma in Western countries, more studies are encouraged to define the microbial profile in these populations.[98]

2.6 CONCLUSION

The primary limitation in microbiome research of oesophagogastric cancer is the lack of standardisation in study methodology. This heterogeneity limits the comparability of data. Nevertheless, trends have been identified in the gastric cancer microbiome with microbial changes towards the progression of benign states to cancer states. There is limited data relating to oesophageal cancer to identify microbial tendencies. The causal link between microbiome and disease, and the correlative relationship between the oral cavity and the tumour tissue has not been fully explored. Furthermore, understanding the functionality and metabolic capabilities of cancer-specific bacteria with a multi-omics approach, could support improved early diagnostics strategies.

Table 4. Specific considerations for designing a microbiome assessment study in upper gastrointestinal disease.

Workflow		Considerations
Study design	Prospective case-control	Cross-sectional observational studies to determine microbial-disease associations Longitudinal studies: premalignant conditions to identify causative factors for diagnostic purposes, therapeutic response, prognostication)
	Patient factors influencing the microbiome	Neoadjuvant treatment naive Use of proton pump inhibitors and/or Histamine-2 antagonists Use of antibiotics (consider time interval between use and microbiome assessment) Immunosuppressive states Synchronous cancer Smoking status Previous gastrointestinal surgery
	Geographical location	Genetic and lifestyle factors such as diet and exercise
	Matching groups	Matched age and gender as a minimum
	Sample size calculation	
	Sub-group analysis	E.g. tumour stage, ethnicity, geographical location and its association with therapeutic response and prognostication – homogeneity will allow a more accurate microbiome assessment
Sampling process	Sample weight	Endoscopic tissue biopsies can be of variable size. Establish a minimum weight of tissue for adequate analysis
	Positive control	Consider adjacent healthy tissue
	Negative control	Consider storing an empty tube and/or storage medium at the same time as the sample
	Replicates	(where possible)
	Minimise freeze-thaw cycles	
Laboratory techniques	Identify potential sources of contamination	Including: inter-laboratory differences, kits and reagents used
	DNA extraction kits	Consider protocols specific to low biomass samples Consider using an empty tube as a negative control
	Batch effect	Minimise batch processing
	PCR	Primer selection and hypervariable region to be targeted
	Sequencing platform	Consider the benefits of whole genome sequencing for metabolic and functional assessment
Recommendation	Data repository	Encourage to deposit reads (raw data and processed data) in an established repository

3. CHAPTER 3: SYSTEMATIC REVIEW: AUGMENTATION OF SHORT CHAIN FATTY ACIDS BY BACTERIAL METABOLISM OF NUTRITIONAL METABOLIC SUBSTRATES

3.1 INTRODUCTION

Short chain fatty acids (SCFA) have been recognised as potential diagnostic biomarkers for oesophagogastric cancer [31-33] but the tumour-associated metabolic network responsible for their production is currently undefined. A hallmark of cancer is the heightened metabolic rate and uncontrolled cell division which may be a source of increasing end-metabolic products. The hypothesis that cancer-related bacteria may be responsible for producing volatile fatty acids warrants further investigation. A clinical pilot study by our group demonstrated the role of oral nutrient stimulants in augmenting target fatty acids within exhaled breath. A 1.6-fold increase of butyric acid was observed 30 minutes after consumption of an oral stimulant drink. Early *in vitro* data supports the clinical study showing an increase in candidate SCFA by exploiting judiciously selected bacterial cell lines treated with glucose or glycerol media: *Escherichia coli* NCIMB 9552, *Klebsiella pneumonia* NCIMB 13281 and *Streptococcus salivarius* NCIMB 701779.

3.2 AIMS

The aim of this systematic review was to:

- (i) Identify nutrient stimulants associated with an increases bacterial short chain fatty acid production.
- (ii) Identify target bacteria associated with short chain fatty acid production.
- (iii) Identify candidate bacterial metabolic pathways responsible for augmentation short chain fatty acid production.

3.3 METHODS

Search strategy

A literature search was performed using electronic databases MEDLINE (OvidSP, 1946 – 03 October 2020) and EMBASE (OvidSP, 1947 – 03 October 2020) to identify articles investigating the augmentation of short chain fatty acids (C2-6). The search terms employed were “(augment* or produc* or enhanc* or amplif* or increas*) adj5 (butanoic acid or butyric acid or pentanoic acid or valeric acid or acetic acid or ethanoic acid or propanoic acid or propionic acid or hexanoic acid or caproic acid) ab,ti.” Proximity of the two search terms was required to narrow the search and retrieve relevant articles.

Studies were included if they reported increased production of a SCFA (C2-C6) in response to a simple metabolic substrate in microbial studies. Exclusion criteria included: (i) provision of a complex nutritional stimuli; (ii) bacteria not known to mammalian species; (iii) articles without original experimental data; (iv) conference abstracts, reviews, letters, notes, comments, and (v) non-English language articles. Three reviewers (BV, KVS, PA) independently screened titles and abstracts in the initial search output. Potentially relevant papers fulfilling the inclusion/exclusion criteria underwent full-text review. Critical data from eligible papers were extracted; (i) target SCFA (ii) nutrient stimulant used (iii) bacteria used (iv) industry/context (v) metabolic pathway considered.

The primary outcome of this review was to identify simple dietary nutrients responsible for SCFA augmentation. Secondary outcomes included identification of mammalian bacteria, and candidate metabolic pathways involved in augmentation.

3.4 RESULTS

A total of 13,783 articles were identified from which relevant 85 studies were selected for analysis. Many excluded studies were due to the use of complex carbohydrate mixtures or use of bacteria not known to be found in mammals. The earliest research identified was by Bhat *et al* in 1947, who augmented butyric acid from lactate fermentation.[99] However, Bhat *et al* describe even earlier research dating back to 1900 where lactate fermentation to produce butyric acid was still in its infancy. Specifically, glucose and lactate were shown to be completely fermented generating higher yields of butyric acid. Similarly, Johns *et al* demonstrated the production of propanoic acid as a result of decarboxylation of succinic acid by rumen bacteria.[100]

Overview:

Substrates The most common group of simple substrates were saccharides, followed by alcohols, polyols, carboxylic acids, and amino acids. Monosaccharides, also known as simple sugars, included glucose, pentose, fructose, galactose, xylose, mannose, arabinose, ribose and rhamnose. They are the most abundant organic compounds in nature, and the simplest form of carbohydrate which does not require further degradation. Disaccharides (sucrose, maltose, 1-kestose, lactose) and trisaccharides (raffinose) were also used. Other energy sources were alcohols (ethanol) and sugar alcohols (galactitol, xylitol, sorbitol, mannitol). **Bacteria** Bacterial cells were sourced from established culture collections or from mammalian species. Five common genera identified were: *Clostridium*, *Lactobacillus*, *Propionibacterium*, *Bifidobacterium*, *Streptococcus*. For some studies, species level of bacteria was used. **Short chain fatty acids** Target SCFA were detected in all cases. Butyric-, propanoic-, and acetic acid were the most common end-products. This biotransformation process is used extensively in the biofuel, pharmaceutical, cosmetic, chemical and food industries to optimise parameters for efficient and maximum yield of SCFA (Table 5). **Glucose:** Glucose is the simplest structural form of a sugar which can immediately enter the glycolytic pathway. It is commonly used to manipulate bacteria and optimise

fermentation parameters to increase the production of butyric-, acetic- and propanoic acids. *Clostridium* is frequently used; particularly *Clostridium tyrobutyricum*, *Clostridium butyricum*, *Clostridium acetobutylicum* and *Clostridium thermoaceticum*. Other bacterial genera include *Streptococcus*, *Propionibacterium*, *Lactobacillus*, *Bacteroides*, and *Bifidobacterium*; all originating from the *Firmicutes* phylum. **Metabolic pathways:** Four of the 85 included studies reported on potential metabolic pathways utilised in the fermentation process. Namdari *et al* observed a glucose-mediated increase of pyruvate dehydrogenase activity, secondary to the partial suppression of the TCA cycle.[101] Abbad-Andaloussi *et al* reported the carbon flux of glycerol and glucose substrates through the fermentation of *Clostridium butyricum* to produce butyrate and acetate. Various branched metabolic pathways may produce these end products. Both substrates can be converted into glyceraldehyde-3-P and subsequently pyruvate. In addition, glycerol can be oxidised and phosphorylated to dihydroxyacetone (DHAP) or dehydrated to 3-hydroxypropionaldehyde (3-HPA). Investigating the carbon intermediates can help define the pathways of metabolism. Lawford *et al* also described the likely contribution of pyruvate catabolism in the production of acetyl-CoA in the process of D-glucose and D-glucuronic acid by *Escherichia coli* B ATCC 11303.[102] Meurice *et al* performed *in vivo* isotope labelled studies with ¹³C-glucose in *Propionibacterium freudenreichii* subsp. *Shermanii*. End metabolic products included labelled- propionate, acetate, succinate and trehalose.[103] Common intermediates and by-products found in both glycolysis and the pentose phosphates pathway were identified: labelled glucose-6-phosphate, glucose-1-phosphate and by-products of carbon dioxide.

The following Table 5 details the nutritional substrates used in a particular bacteria fermentation process to produce higher yields of SCFA. The context of each experiment has been described and is important to demonstrate that this process is feasible and has been optimised in large industries.

Table 5. Short chain fatty acid augmentation by bacterial metabolism using nutritional stimuli.

SUBSTRATE		GENERA/SPECIES	METABOLIC END PRODUCT	CONTEXT (INDUSTRY)
MONOSACCHARIDE				
Glucose	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium tyrobutyricum, Clostridium butyricum, Clostridium acetobutylicum, Clostridium aminovalericum, Clostridium cochlearium, Clostridium tetani, Clostridium sporosphaeroides, Clostridium thermoaceticum (ATCC 39073)	Butyric acid, Acetic acid, Propanoic acid	Biofuel, Food, Chemical, Pharmaceutical, Cosmetic [104-114]
	P: Firmicutes, C: Negativicutes, O: Seleomonadales, F: Selenomonadaceae, G: Selenomonas	Selenomonas bovis (Yak rumen)		<i>In vitro</i> [115-118]
	P: Firmicutes, C: Clostridia, O: Thermoanaerobacterales, F: Thermoanaerobacteraceae, G: Tepidanaerobacter	Tepidanaerobacter		
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium thermoaceticum	Acetic acid	Biofuel, Food, Cosmetic [119] <i>In vitro</i> [120]
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Streptococcaceae, G: Aeromonas	viz Streptococcus gordonii, Streptococcus mutans, Streptococcus oralis, Streptococcus salivarius, Streptococcus sanguis and Streptococcus sobrinus		<i>In vitro</i> (cell lines) [121]
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Staphylococcaceae, G: Staphylococcus	Staphylococcus carnosus		Food [122]
	P: Proteobacteria, C: Gammaproteobacteria, O: Aeromonadales, F: Aeromonadaceae, G: Streptococcus	Aeromonas strains		Biofuel [101]
	P: Proteobacteria, C: Gammaproteobacteria, O: Enterobacterales, F: Enterobacteriaceae, G: Escherichia	Escherichia coli B		Biofuel [102] <i>In vitro</i> (cell lines) [123]
	P: Proteobacteria, C: Alphaproteobacteria, O: Rhodospirillales, F: Acetobacteraceae, G: Acetobacter	Acetobacter sp. (Drosophila melanogaster)		
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus coagulans (probiotic)	Propanoic acid	Food, Chemical, Pharmaceuticals [124]
	P: Firmicutes, C: Negativicutes, O: Veillonellales, F: Veillonellaceae, G: Veillonella	Veillonella criceti		
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus sp.	Acetic acid, Propanoic acid	Biofuel, Chemical [111] Food [125] <i>In vitro</i> [126-128]
	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici, Propionibacterium freudenreichii subsp. Shermanii		Biofuel, Food, Cosmetic, Pharmaceuticals [129] <i>In vitro</i> [118, 128]

	P: Bacteroidetes, C: Bacteroidia, O: Bacteroidales, F: Bacteroidaceae, G: Bacteroides	Bacteroides fragilis		Biofuel, Food, Chemical, Pharmaceuticals, Cosmetics [130-137] <i>In vitro</i> (cell lines) [103, 118, 138] Biofuel, Food, Chemical [139]
	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacteria, Bifidobacterium animalis DN-173 010, Bifidobacterium longum NCC2705	SCFA, Acetic acid	<i>In vitro</i> (human intestinal) [140] Food [141, 142]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Peptostreptococcaceae, G: Peptostreptococcus	Peptostreptococcus	Butyric acid, Acetic acid	<i>In vitro</i> [117]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Eubacteriaceae, G: Eubacterium	Eubacterium	Butyric acid, Propanoic acid, Pentanoic acid	<i>In vitro</i> [118]
Hexose/Pentose	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium tyrobutyricum	Butyric acid	Biofuel, Food, Chemical, Pharmaceuticals [143]
Fructose	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus	Acetic acid	Food [125, 144] <i>In vitro</i> [126]
	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacterium animalis DN-173 010		Food [141]
Galactose	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium tyrobutyricum	Butyric acid	Biofuel [105, 106] Food, Chemical, Pharmaceuticals [105]
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus intermedius NRRL B-3693	Acetic acid	<i>In vitro</i> [126]
	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacterium animalis DN-173 010		Food [141]
Xylose	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium tyrobutyricum, Clostridium beijerinckii	Butyric acid	Biofuel, Chemical, Pharmaceuticals, Food [104, 145]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium	Butyric acid, Acetic acid	Biofuel, Chemical [111, 146]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Eubacteriaceae, G: Eubacterium	Eubacterium		
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus coagulans (probiotic)	Propanoic acid, Acetic acid	Food, Pharmaceuticals, Chemical, Biofuel [111, 124]
	P: Firmicutes, C: Negativicutes, O: Veillonellales, F: Veillonellaceae, G: Veillonella	Veillonella criceti		
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus plantarum	Acetic acid	Food [147]
	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium shermanii	Propanoic acid	<i>In vitro</i> [128]
		Lactobacillus xylosus		

	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus			
Mannose	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Bifidobacterium longum NCC2705 Lactobacillus intermedius NRRL B-3693	Acetic acid	<i>In vitro</i> [126] Food [142]
Arabinose	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus plantarum	Acetic acid	Food [147]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium acetobutylicum, Clostridium butylicum	Acetic acid, Butyric acid, Propanoic acid	<i>In vitro</i> [116]
Ribose	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus plantarum	Acetic acid	Food [147]
Rhamnose	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium acetobutylicum, Clostridium butylicum	Acetic acid, Butyric acid, Propanoic acid	<i>In vitro</i> [116]
DISACCHARIDE				
Sucrose	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium P: Bacteroidetes, C: Bacteroida, O: Bacteroidales, F: Bacteroidaceae, G: Bacteroides P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium P: Firmicutes, C: Clostridia, O: Clostridiales, F: Eubacteriaceae, G: Eubacterium	Bifidobacteria Bacteroides Clostridium Eubacterium	Acetic acid, Butyric acid, Propanoic acid	Food [125, 141] <i>In vitro</i> (human faeces), Food [148]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Ruminococcaceae, G: Ruminococcus P: Firmicutes, C: Negativicutes, O: Seleomondales, F: Veillonellaceae, G: Megasphaera	Ruminococcaceae bacterium CPB6 Megasphaera elsdenii NCIMB 702410	Hexanoic acid	Biofuel, Chemical, Food [149, 150]
	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici ATCC 4875	Propanoic acid	Biofuel, Cosmetic, Food, Chemical, Pharmaceuticals [130, 151]
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus intermedius NRRL B-3693	Acetic acid	<i>In vitro</i> [126]
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Bacillus strain Clostridium tyrobutyricum ATCC 25755	Butyric acid	Biofuel, Food, Chemical, Pharmaceuticals [152]
Maltose	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus	Acetic acid	Food [125, 153] <i>In vitro</i> [126]

	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici	Propanoic acid	Biofuel, Food, Chemical [154]
1-kestose (fructose+sucrose)	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacteria	Butyric acid	Food, <i>In vivo</i> (caecum cats) [155]
Lactose	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacteria, Bifidobacterium animalis DN-173 010	Short chain fatty acids, Acetic acid, Propanoic acid	<i>In vitro</i> (human faeces) [156] Food [141, 156, 157]
	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus helveticus ATCC 15807	Propanoic acid, Acetic acid	Food [158, 159]
	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici, Propionibacterium freudenreichii		Biofuel, Cosmetics, Chemical, Food, Pharmaceuticals [135, 154, 157, 160-162]
	P: Proteobacteria, C: Gammaproteobacteria, O: Enterobacteriales, F: Enterobacteriaceae, G: Salmonella	Salmonella typhimurium	Propanoic acid	<i>In vitro</i> (caecum broiler chickens) [163]
	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium shermanii		Food industry [159]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium kluyveri	Hexanoic acid	Biofuel, Chemical [164]
TRISACCHARIDE				
Raffinose	P: Firmicutes, C: Bacilli, O: Lactobacillales, F: Lactobacillaceae, G: Lactobacillus	Lactobacillus intermedius NRRL B-3693	Acetic acid	<i>In vitro</i> [126]
ALCOHOLS				
Ethanol	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium kluyveri	Hexanoic acid, Butyric acid	Biofuel, Chemical, Food [164, 165]
	P: Fibrobacteres, C: Fibrobacteria, O: Fibrobacterales, Fa: Fibrobacteraceae, G: Fibrobacter	Fibrobacter succinogenes S85 (cattle)		Biofuel, Food, Chemical [165]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Ruminococcaceae, G: Ruminococcus	Ruminococcus		Biofuel, Food, Chemical [165]
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus thermoamylovorans	Acetic acid, Butyric acid, Propanoic acid	Biofuel [166]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium aminovalericum, Clostridium cochlearium, Clostridium tetani, Clostridium sporosphaeroides		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Veillonellaceae, G: Dendrosporobacter	Dendrosporobacter querciculus		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Selenomonadaceae, G: Selenomonas	Selenomonas bovis		

	P: Firmicutes, C: Clostridia, O: Thermoanaerobacterales, F: Thermoanaerobacteraceae, G: Tepidanaerobacter	Tepidanaerobacter		
SUGAR ALCOHOL				
Galactitol	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium	Butyric acid, Acetic acid, Hexanoic acid	Biofuel, Chemical [167]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Ruminococcaceae, G: Caproiciproducens	Caproiciproducens galactitolivorans	Hexanoic acid	Biofuel, Chemical [167, 168]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium sp. BS-1		Biofuel, Food, Chemical, Pharmaceuticals, Cosmetics [169]
Xylitol	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium, Phascolarctobacterium, Ecoli, Bifidobacterium	Butyric acid	<i>In vitro</i> (human faeces), Food, Chemical, Pharmaceuticals [170]
Sorbitol	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici	Propanoic acid	Biofuel, Food, Chemical, Pharmaceuticals [171]
Mannitol	P: Firmicutes, C: Cocci, O: Lactobacillales, F: Streptococcaceae, G: Streptococcus	Viridans streptococci	Butyric acid	<i>In vitro</i> (caecum rat) [172]
	P: Actinobacteria, C: Actinobacteria, O: Bifidobacteriales, F: Bifidobacteriaceae, G: Bifidobacterium	Bifidobacteria		
SIMPLE POLYOL				
Glycerol	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium freudenreichii	Propanoic acid, Acetic acid	Biofuel, Food, Chemical, Pharmaceuticals, Cosmetics [133, 139, 173-175]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium, Clostridium butyricum DSM 5431	Propanoic acid, Butyric acid	Biofuel, Food, Chemical, Pharmaceuticals, Cosmetics [113, 176]
	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus thermoamylovorans	Acetic acid, Butyric acid, Propanoic acid	Biofuel, Chemical [111, 166] <i>In vitro</i> [129]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium aminovalericum, Clostridium cochlearium, Clostridium tetani, Clostridium sporosphaeroides		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Veillonellaceae, G: Dendrosporobacter	Dendrosporobacter quercicolus		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Selenomonadaceae, G: Selenomonas	Selenomonas bovis		
	P: Firmicutes, C: Clostridia, O: Thermoanaerobacterales, F: Thermoanaerobacteraceae, G: Tepidanaerobacter	Tepidanaerobacter		

	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici, Propionibacterium acnes, Propionibacteriumthoenii (NCDO 1082)	Propanoic acid	<i>In vitro</i> [129] Biofuel, Food, Chemical, Pharmaceuticals, Cosmetics [131, 133, 176-182]
MONOCARBOXYLIC ACID				
Lactate	P: Actinobacteria, C: Actinobacteria, O: Propionibacteriales, F: Propionibacteriaceae, G: Propionibacterium	Propionibacterium acidipropionici	Propanoic acid, Acetic acid	Biofuel, Pharmaceuticals, Cosmetics, Food, Chemical [135, 177] <i>In vivo</i> (ruminal sheep) [100]
	P: Firmicutes, C: Negativicutes, O: Veillonellales, F: Veillonellaceae, G: Veillonella	Veillonella gazogenes		
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium lacto-acetophilum nov. spec	Butyric acid	<i>In vitro</i> [99]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium kluyveri	Hexanoic acid	Biofuel, Chemical [164]
Acetate	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium lacto-acetophilum nov. spec	Butyric acid	<i>In vitro</i> [99]
DICARBOXYLIC ACID				
Succinate	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus thermoamylovorans	Acetic acid, Butyric acid, Propanoic acid	Biofuel [166]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium aminovalericum, Clostridium cochlearium, Clostridium tetani, Clostridium sporosphaeroides		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Veillonellaceae, G: Dendrosporobacter	Dendrosporobacter querciculus		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Selenomonadaceae, G: Selenomonas	Selenomonas bovis		
	P: Firmicutes, C: Clostridia, O: Thermoanaerobacterales, F: Thermoanaerobacteraceae, G: Tepidanaerobacter	Tepidanaerobacter		
	P: Firmicutes, C: Negativicutes, O: Veillonellales, F: Veillonellaceae, G: Veillonella	Veillonella gazogenes	Acetic acid, Propanoic acid	<i>In vivo</i> (ruminal sheep) [100]
OTHER				
Amino acids	P: Firmicutes, C: Bacilli, O: Bacillales, F: Bacillaceae, G: Bacillus	Bacillus thermoamylovorans	Acetic acid, Butyric acid, Propanoic acid	Biofuel [166]
	P: Firmicutes, C: Clostridia, O: Clostridiales, F: Clostridiaceae, G: Clostridium	Clostridium aminovalericum, Clostridium cochlearium, Clostridium tetani, Clostridium sporosphaeroides		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Veillonellaceae, G: Dendrosporobacter	Dendrosporobacter querciculus		
	P: Firmicutes, C: Negativicutes, O: Selenomonadales, F: Selenomonadaceae, G: Selenomonas	Selenomonas bovis		

	P: Firmicutes, C: Clostridia, O: Thermoanaerobacterales, F: Thermoanaerobacteraceae, G: Tepidanaerobacter	Tepidanaerobacter		
Sodium gluconate / glucuronic acid	P: Proteobacteria, C: Gammaproteobacteria, O: Enterobacterales, F: Enterobacteriaceae, G: Escherichia	Escherichia coli strain KO11, Escherichia coli B (ATCC 11303)	Acetic acid	Biofuel industry [102, 183]

3.5 DISCUSSION

The principal findings of this review were that target SCFA, namely butyric-, acetic- and propanoic acids, can be augmented by exposing select bacteria to nutrient stimuli. This biotransformation process has been of interest since the 1900s to optimise the production of these fatty acids for industry.[99] Few studies have explored the metabolic pathways involved in conversion of substrates to end metabolic products. However, four studies have alluded to the contribution of multiple pathways, although the action of one may be predominant. There is a common pathway for glycerol and glucose metabolism forming G3P, in addition to independent pathways. Glycerol has been linked to the DHAP pathway, and glucose labelled studies have identified important intermediates via the G6P and pentose phosphates pathways. Selection of bacteria has also proven to be pivotal in how substrates are metabolised and ultimately the resulting end-products. *Clostridium*, for example, is considered a “butyric acid fermenter.” The majority of papers used *Clostridium* species as a source of butyric acid production, with common by-products including acetic- and propanoic acids. Glucose was the most selected substrate to generate a higher SCFA yield. It is the simplest form of a carbohydrate, which does not require a degradation process before entering the metabolic pathway for fatty acid production. This translates to efficiency and increased yield in industry.

3.6 CONCLUSION

This review demonstrates a proof of concept that augmentation of SCFA is feasible by feeding simple substrates to select bacteria. This is not limited to carbohydrates, but also include amino acids and carboxylic acids. Optimal environmental conditions are required, but this review supports that the pathways enabling this process, do exist, and are commonly used in industry. The key is judiciously selecting the optimal combination of simple substrate and bacteria to produce target end-products.

SECTION 2

CHARACTERISATION OF THE OESOPHAGOGASTRIC CANCER ASSOCIATED MICROBIOME

4. CHAPTER 4: CHARACTERISATION OF OESOPHAGOGASTRIC CANCER MICROBIOME USING 16S rRNA GENE SEQUENCING

4.1 INTRODUCTION

The oesophagogastric microbiome has gained an increasing interest in the last 10 years. The upper gastrointestinal tract harbours bacteria at 10^1 per gram of contents, which is a fraction compared to the distal gut at 10^{12} per gram of contents.[184] Thus far, the core microbial distribution comprises five phyla: *Firmicutes* (83%), *Proteobacteria* (8%), *Bacteroidetes* (7%), *Actinobacteria* (2%), and *Fusobacteria* (1%). Similarities have been identified in the microbiome of the oral cavity (10^{11} per gram of contents) and the oesophagus suggesting there is a translocation of bacteria during swallowing which may alter the oesophageal microenvironment.[185] Although there is limited experimental data, associations between bacteria and oesophageal cancer have been made. An increasing abundance of *Fusobacterium nucleatum* has been correlated with oesophageal squamous cell carcinoma, translating to significantly reduced survival outcomes. The stomach presents a unique acidic environment contributing to a fastidious ecological network. *Helicobacter pylori* is known to induce gastritis and peptic ulcer disease and has been linked to gastric cancer in 3% of infected patients. However, the understanding of the oesophagogastric cancer microbiome is still in its infancy and warrants further investigation with the advancing and evolving sequencing technologies now available.

4.2 AIMS

The primary aim of this study is to characterise the oral and tumour-associated microbiome of oesophagogastric adenocarcinoma compared to healthy controls.

4.3 METHODS

Patient selection Patients with histologically confirmed oesophageal (n=12) or gastric adenocarcinoma (n=12), were recruited prior to surgery. Age-matched healthy controls undergoing endoscopic investigation for upper gastrointestinal symptoms were recruited from the endoscopy department (n=27). All patients were recruited from Imperial College Healthcare Trusts (St Mary's Hospital and Hammersmith Hospital) between March 2019 and October 2020. Exclusion criteria were: synchronous cancer diagnosis at another site, immunosuppression, severe systemic disease, recent antibiotic use within 2 weeks. All patients were provided with patient information leaflets, and written informed consent was obtained. Ethical approval was obtained via the Imperial College Healthcare Tissue Bank; REC Wales approval: 17/WA/0161 (R20059).

Sample preparation, collection, and storage Stocks of 50:50 99% glycerol and 0.9% normal saline were prepared in cryovials and degassed with pure nitrogen to create an anaerobic environment. Literature findings have reported the use of 10-50% glycerol dilution to prevent osmotic shock, resulting in bacterial cell wall damage and cell death. At the correct dilution, glycerol fuses into the bacterial cell wall to maintain structural integrity and secure cell viability at -80°C storage. Four tissue biopsies were taken per site from the resection specimen using sterile forceps and scalpel blade: (i) normal oesophagus (ii) normal stomach and (iii) tumour. Endoscopic samples (up to 4mm in size) were taken for the healthy control cohort and oesophagogastric cancer patients undergoing a staging laparoscopy with cold biopsy forceps (2.8mm, EndoJaw, Olympus, UK). All samples were snap frozen in a dry ice-pure ethanol bath, and immediately transferred to (i) a glycerol stock and (ii) empty cryovial. Samples were stored at -80°C until further processing.

Bacterial 16S rRNA gene extraction All samples were thawed on ice, and DNA extracted using the DNeasy® PowerLyzer® PowerSoil® Kit (Qiagen). The tissue was added to the PowerBead Tube with 750µL of the PowerBead solution and 60µL of Solution C1 before homogenising the sample for 10 minutes with a vortex adapter for chemical and mechanical lysis of the cell wall. Following centrifugation at 10,000 g for 30 seconds, the supernatant was transferred to a clean tube, and 250µL of Solution C2 was added, incubated at 4°C for 5 minutes and centrifuged at 10,000 g for 1 minute. The supernatant was aliquoted, 200µL of Solution C3 was added before incubation for 5 minutes at 4°C and centrifuged at 10,000 g for 1 minute. After adding 1200µL to the supernatant, 675µL was centrifuged using the MB spin column at 10,000 g for 1 minute. This was repeated to ensure purification of the sample. 500µL of an ethanol solution (C5) was added, centrifuged for 30 seconds and repeated again. The spin filter was carefully placed into a clean tube with 100µL Solution C6 (10mM Tris) and centrifuged at 10,000 g for 30 seconds to elute the DNA from the spin filter membrane. All reagents used precipitated non-DNA material to isolate pure DNA.

Polymerase Chain Reaction (PCR): First Stage

Variable regions V3-V4 were targeted for amplification using the 16S Metagenomic Sequencing Library Preparation protocol for Illumina MiSeq. DNA quantification was performed using the Qubit™ dsDNA high-sensitivity assay (Invitrogen™). A concentration of 5ng/µL was achieved for each sample. Samples with too low concentration were evaporated on the SpeedVac Concentrator (Savant RVT4104) and samples with too high concentration were diluted up to 1:10 to achieve a final concentration of 5ng/µL. A 25µL reaction mixture comprising 2.5µL genomic DNA, 5µL reverse primer (1µM), 5µL forward primer (1µM) and 12.5µL 2x KAPA HiFi HotStart ReadyMix (Roche®) was set-up in a 96-well plate and placed in a thermal cycler (Veriti™ Applied Biosystems™) at the following conditions:

- Initiation: 95°C for 3 minutes
- 25 cycles of:
 - Denaturation to dehybridize the DNA strands by breaking hydrogen bonds: 95°C for 30 seconds

- Annealing to allow primers to base pair with complementary single strand DNA templates: 55°C for 30 seconds
- Elongation: 72°C for 30 seconds
- Final extension to activate DNA polymerase and begin adding nucleotides to the primer: 72°C for 5 minutes
- Hold at 4°C

Table 6. 16S rRNA gene amplicon PCR primers.

16S gene amplicon PCR Primers	
Forward Primer	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
Reverse Primer	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

Samples were transferred to 0.2mL 8-microAmp tube strips and purified with AMPure XP beads to isolate variable regions V3-V4 away from the primers. 20µL of beads were added, gently mixed, and incubated at room temperature for 5 minutes. The tube was placed onto a magnetic stand for 2 minutes until the supernatant was clear and subsequently discarded. The beads were washed with 200µL of 80% ethanol, incubated for 30 seconds on the magnetic stand and supernatant discarded. A second ethanol wash was performed, after which the sample was air-dried for 10 minutes. The tube strip was removed from the magnetic stand and beads resuspended in 52.5µL 10mM Tris pH8.5 elution buffer. After 2 minutes of incubation, a further 2 minutes was completed in the magnetic stand until the supernatant was clear. 50µL of the supernatant was harvested for downstream processing.

1µL of the amplicon PCR product was analysed with High Sensitivity Agilent D1000 TapeStation to assess successful amplification of V3-V4 region; 550 base pairs (bp). D1000 allows for the analysis of double stranded DNA molecules between 35 and 1000 base pairs. However, the trace did not demonstrate a peak at 550bp. The previous steps were reviewed, and the likely cause was the evaporation of the 25µL mixture during thermal cycling. A residual 18µL was used for the clean-up process. To address this, 8-microAmp tube strips with caps (Applied Biosystems™) were used instead of PCR plates to ensure effective sealing to prevent sample evaporation. In addition, 26µL of the eluent buffer was used to increase DNA concentration yield. First stage PCR and clean-up was repeated in this fashion and validated with identifiable 550bp using the D100 TapeStation (Figure 3).

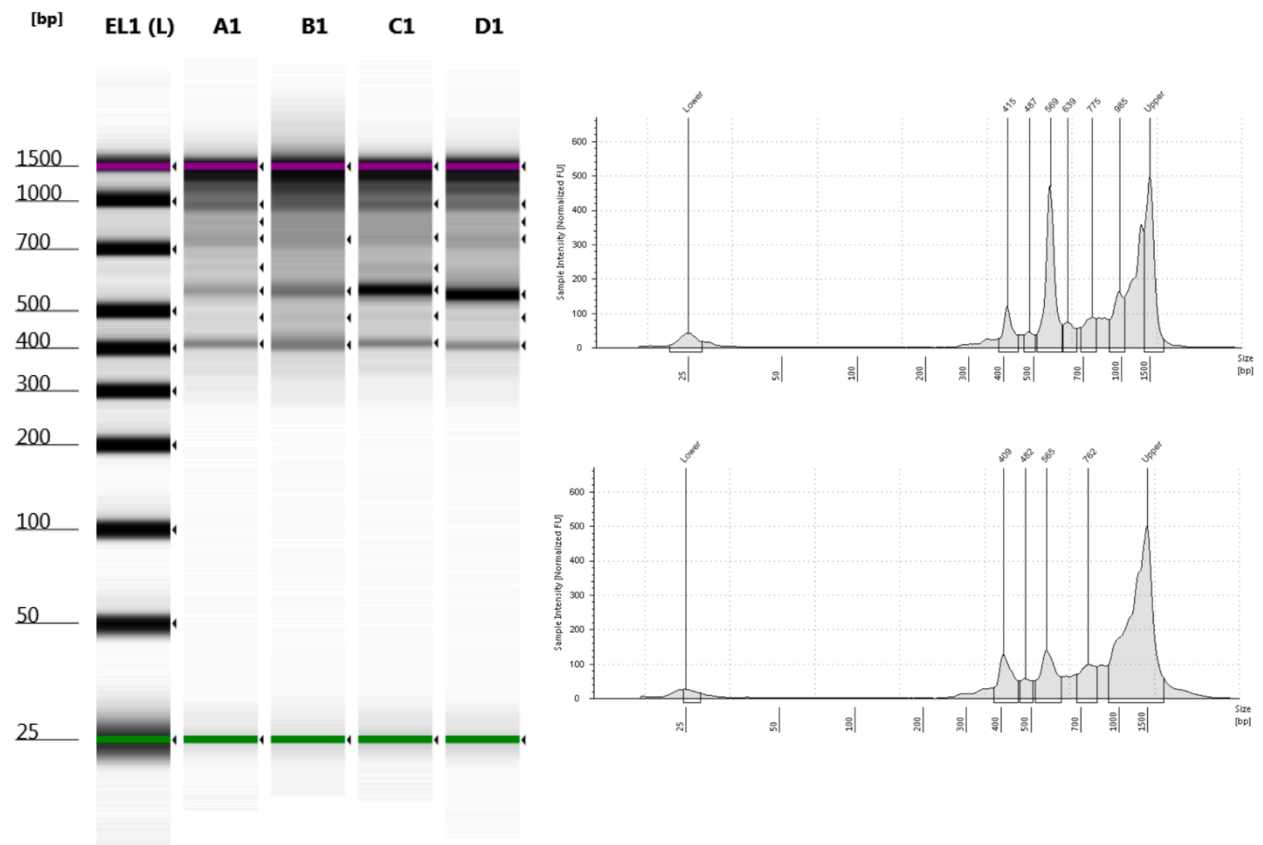


Figure 3. D1000 TapeStation trace demonstrating 16S rRNA gene amplicon PCR band present at ~550bp for V3-V4 regions. EL1 (L) = ladder. A1 = blank. B1= blank. C1 = Sample 1. D1 = Sample 2.

Polymerase Chain Reaction (PCR): Second Stage

Second stage PCR was performed using the Nextera XT Index Kit. 5µL of the amplicon product was mixed with the Index 1 (N7XX) and 2 (S5XX) primers, 2x KAPA HiFi HotStart ReadyMix (Roche®) and PCR grade water. The reaction mixture was placed in a thermal cycler (Veriti™ Applied Biosystems™) with the following conditions:

- 95°C for 3 minutes
- 8 cycles of:
 - 95°C for 30 seconds
 - 55°C for 30 seconds
 - 72°C for 30 seconds
- 72°C for 5 minutes
- Hold at 4°C

PCR clean-up was performed as described above with a resulting 25µL of supernatant from the Index PCR plate. Validation was performed using the DNA D1000 chip to confirm a trace at 630bp. Samples underwent sequencing on the MiSeq platform.

Microbiome and statistical analysis

Raw data files (.fastq) were analysed using the DADA2 pipeline on R statistical package (version 4.1.0). Forward and reverse reads were quality assessed to trim and filter reads falling below Phred score 30. Forward and reverse reads were truncated at 230- and 160 base pairs respectively. Microbiome analyst was used to calculate alpha diversities (Chao1, ACE, Shannon, Simpson, Fisher) with non-parametric Mann Whitney U test analysis, beta diversities (Bray-Curtis, Jensen-Shannon, Jaccard Index, Un/weighted Unifrac) examined with pairwise PERMANOVAs, phylogenetic tree and univariate non-parametric analyses. Bacteria known to be contaminants in DNA extraction kits were excluded. Linear discriminant analysis Effect size was performed (Galaxy; huttenhower.sph.harvard.edu) with a Kruskal Wallis analysis using p value < 0.05 as statistically significant, and a logarithmic LDA score of 2.0.

4.4 RESULTS

Gastric adenocarcinoma No significant alpha or beta diversity measures were identified at any taxonomic level (Table 7). Shannon and Simpson indices showed visual trends towards reduced alpha diversity in bacteria of gastric cancer tissue from phyla to genera (Figure 4). Significant beta diversity was demonstrated between the normal stomach of control patients and normal stomach adjacent to the gastric cancer for all 5 metrics and across all 5 taxonomies; $p < 0.05$ (Figure 5).

Comparisons were made between 3 bacteria groups: (i) gastric cancer vs control normal stomach (ii) gastric cancer vs normal stomach adjacent to gastric cancer and (iii) normal stomach adjacent to gastric cancer vs control normal stomach. A higher abundance of *Lactobacillus* from the family, genus and species levels was observed in gastric cancer tissue compared to normal stomach from controls; *Lactobacillaceae* ($p=0.04$), *Lactobacillus* ($p=0.04$) and *Lactobacillus vaginalis* ($p=0.006$). Similarly, *Pseudopropionibacterium* ($p=0.047$) and its species *Pseudopropionibacterium propionicum* ($p=0.047$) were highly abundant in gastric cancer tissue. Significant differences were observed between normal stomach adjacent to gastric cancer and normal stomach from healthy control patients. *Prevotellaceae* ($p=0.04$) and *Prevotella* ($p=0.048$) were less abundant in the normal stomach of gastric cancer patients. However, *Alphaproteobacteria* ($p=0.02$), *Achromobacter* ($p=0.01$) and *Lactobacillus vaginalis* ($p=0.01$) were more prevalent in normal stomach adjacent to cancer tissue (Figure 6, Table 8). LEfSe analysis showed an abundance of *Solobacterium* genus and 9 species including *Lactobacillus delbrueckii*, *Streptococcus lactarius* and *Prevotella pleuriditis* (Figure 7).

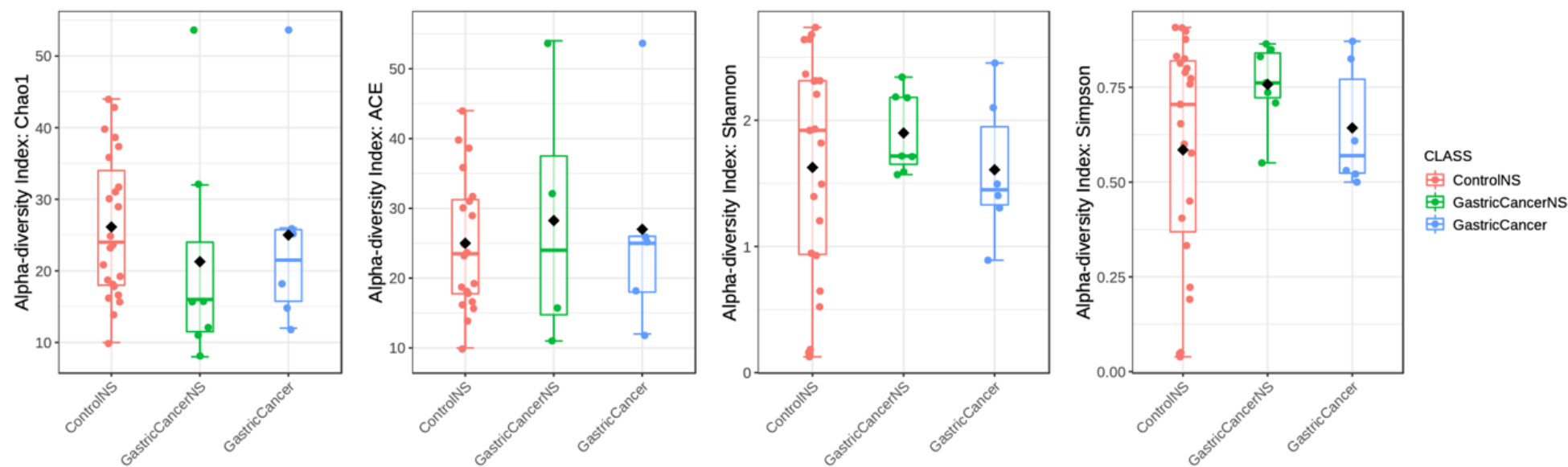


Figure 4. Alpha diversity metrics of bacteria genera comparing Control normal stomach (red) to Gastric cancer adjacent normal stomach (green) and Gastric cancer (blue); $p > 0.05$. No significant differences were identified.

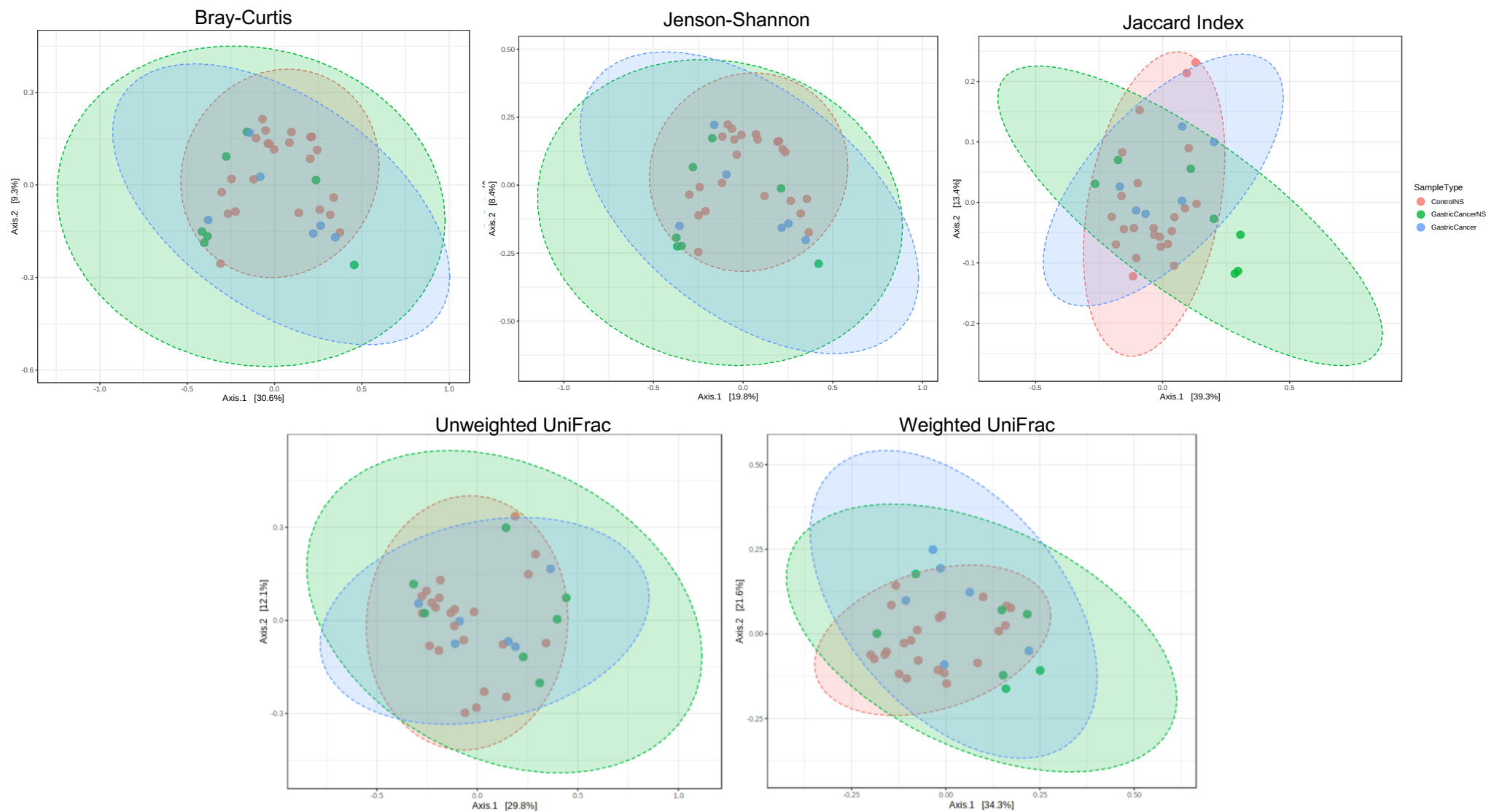


Figure 5. Beta diversity metrics of bacteria genera comparing Control normal stomach (red) to Gastric cancer adjacent normal stomach (green) and Gastric cancer (blue); Bray-curtis $p=0.004$, Jenson-shannon $p=0.03$, Jaccard index $p=0.001$, Unweighted Unifrac $p=0.02$, Weight Unifrac $p=0.02$.

Table 7. Alpha and beta diversities of bacterial communities in gastric cancer tissue, adjacent normal stomach, and normal stomach from healthy controls. NS = normal stomach. P value < 0.05 is considered statistically significant (highlighted in red).

	Alpha diversity					Beta diversity				
	Chao 1	ACE	Shannon	Simpson	Fisher	Bray-Curtis	Jensen-Shannon	Jaccard Index	Unweighted Unifrac	Weighted Unifrac
Gastric cancer vs Control NS										
Phylum	0.70	-	0.66	0.81	-	0.15	0.25	0.18	0.60	0.09
Class	0.74	0.37	0.90	0.81	0.38	0.08	0.30	0.10	0.76	0.07
Order	0.83	0.59	0.66	1.00	0.30	0.22	0.32	0.19	0.62	0.11
Family	0.65	0.81	0.73	0.94	0.23	0.28	0.20	0.22	0.52	0.09
Genus	0.63	1.00	0.85	0.90	0.21	0.26	0.22	0.20	0.52	0.09
Species	0.89	0.95	1.00	0.85	0.55	0.33	0.27	0.23	0.52	0.09
Gastric cancer vs Gastric cancer NS										
Phylum	0.72	-	0.73	0.63	-	0.05	0.07	0.07	0.94	0.11
Class	0.35	0.64	0.95	0.53	-	0.08	0.11	0.11	0.32	0.12
Order	0.22	0.63	0.37	0.29	0.37	0.10	0.36	0.11	0.54	0.17
Family	0.29	0.90	0.53	0.23	0.53	0.26	0.36	0.24	0.45	0.29
Genus	0.35	0.90	0.23	0.23	0.53	0.23	0.36	0.24	0.44	0.20
Species	0.32	1.00	0.53	0.63	0.37	0.28	0.39	0.25	0.39	0.29
Gastric cancer NS vs Control NS										
Phylum	0.33	1.00	0.74	0.77	-	0.01	0.02	0.02	0.41	0.03
Class	0.30	0.18	0.89	0.89	0.29	0.02	0.02	0.02	0.04	0.02
Order	0.18	0.67	0.63	0.36	0.21	<0.01	0.01	<0.01	0.01	0.01
Family	0.19	0.83	0.63	0.44	0.19	0.01	0.02	<0.01	0.01	0.01
Genus	0.12	1.00	0.67	0.33	0.14	<0.01	0.03	<0.01	0.02	0.02
Species	0.12	0.92	0.81	0.33	0.19	0.01	0.05	<0.01	0.02	0.02

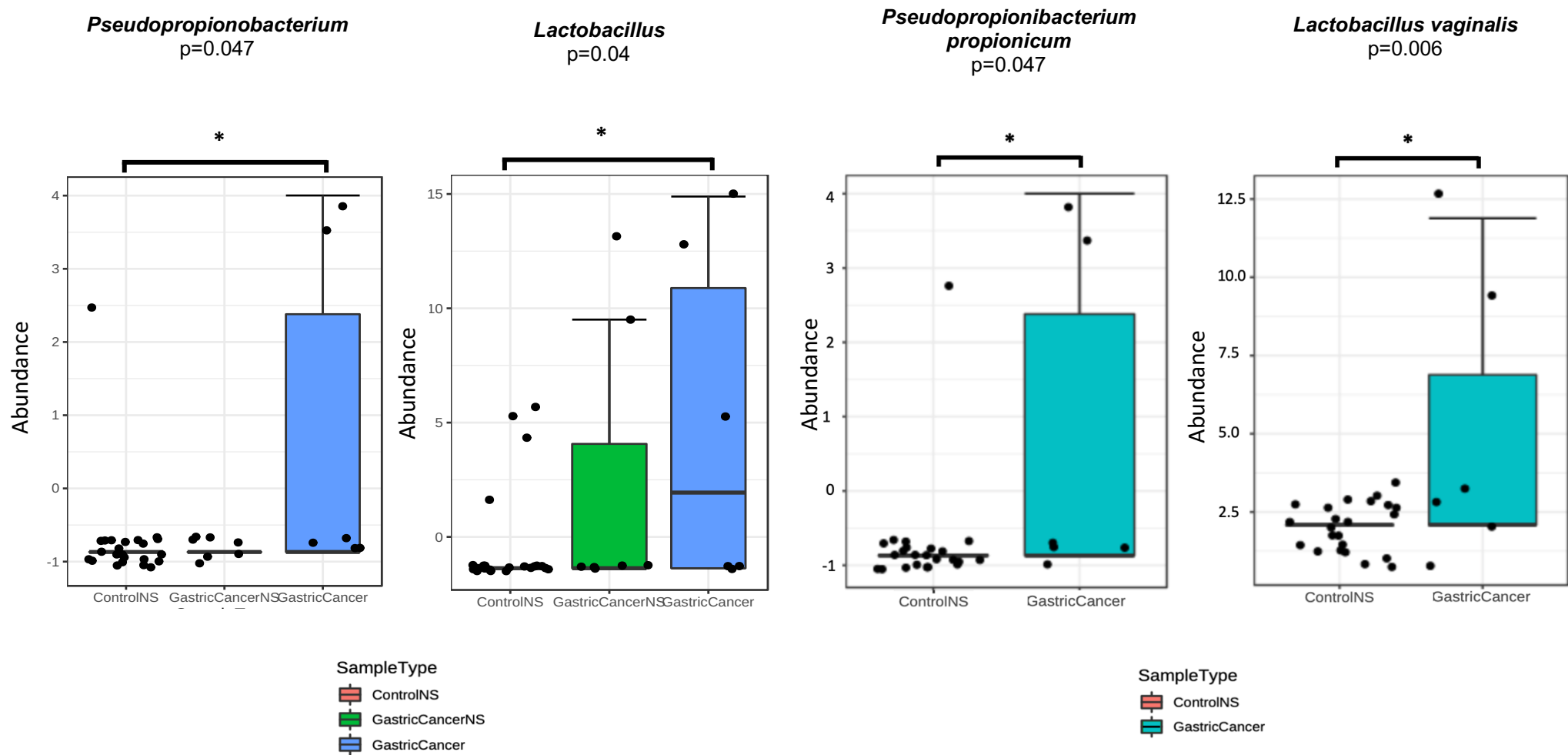


Figure 6. Box plots demonstrating select bacteria in higher abundances in gastric cancer tissue compared to healthy control normal stomach. P value < 0.05 is considered statistically significant.

Table 8. Bacteria significantly altered in gastric cancer-specific bacteria. ↑ = high abundance. ↓ = low abundance.

	P value	In cancer-related bacteria
Gastric cancer vs Control NS		
<i>Bacilli</i> (Class)	0.047	↑
<i>Lactobacillaceae</i> (Family)	0.04	↑
<i>Pseudopropionibacterium</i> (Genus)	0.047	↑
<i>Lactobacillus</i> (Genus)	0.04	↑
<i>Actinomyces naeslundii</i> (Species)	0.006	↑
<i>Lactobacillus vaginalis</i> (Species)	0.006	↑
<i>Oribacterium parvum</i> (Species)	0.01	↑
<i>Granulicatella elegans</i> (Species)	0.025	↑
<i>Pseudopropionibacterium propionicum</i> (Species)	0.047	↑
<i>Lachnoanaerobaculum cf</i> (Species)	0.047	↑
Gastric cancer vs Gastric cancer NS		
<i>Lactobacillales</i> (Order)	0.03	↑
Gastric cancer NS vs Control NS		
<i>Bacteroidales</i> (Order)	0.04	↓
<i>Alcaligenaceae</i> (Family)	0.01	↑
<i>Beijerinckiaceae</i> (Family)	0.01	↑
<i>Prevotellaceae</i> (Family)	0.04	↓
<i>Yersiniaceae</i> (Family)	0.04	↑
<i>Achromobacter</i> (Genus)	0.01	↑
<i>Serratia</i> (Genus)	0.001	↑
<i>Prevotella</i> (Genus)	0.048	↓
<i>Methylobacterium Methylobacterium</i> (Genus)	0.01	↑
<i>Bifidobacterium breve</i> (Species)	0.01	↑
<i>Haemophilus haemolyticus</i> (Species)	0.01	↑
<i>Neisseria mucosa</i> (Species)	0.01	↑
<i>Lactobacillus vaginalis</i> (Species)	0.01	↑
<i>Leptotrichia wadei</i> (Species)	0.01	↑

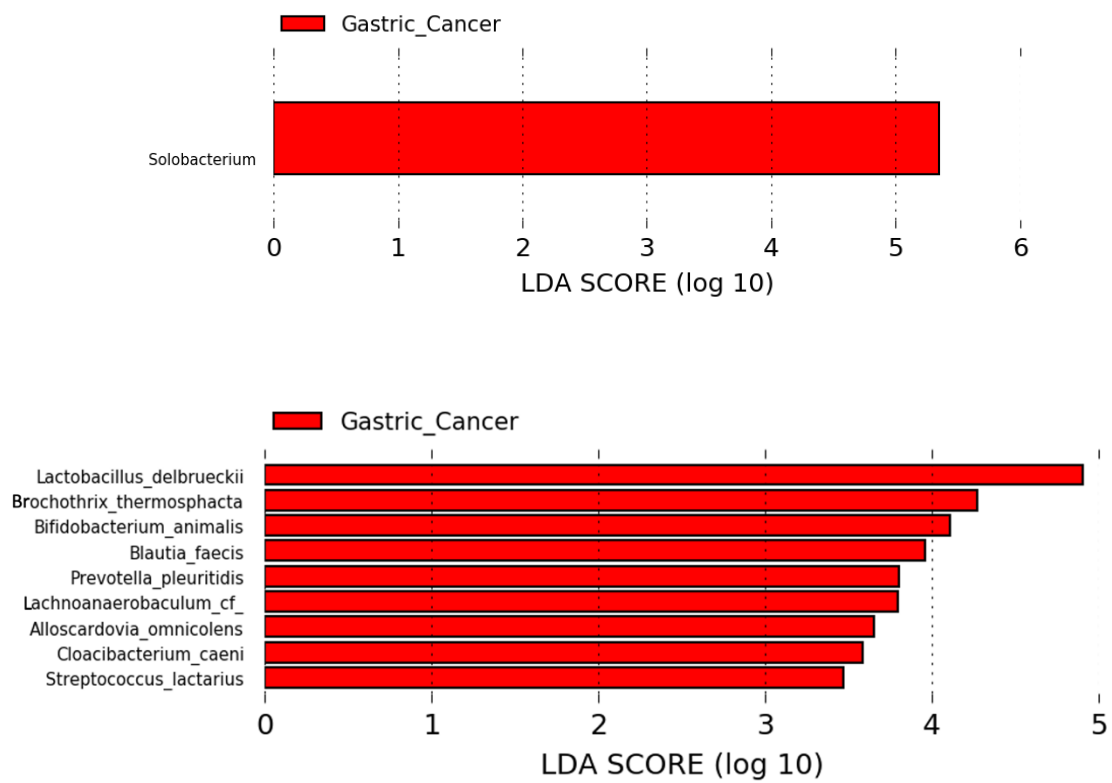


Figure 7. Linear discriminant analysis Effect Size demonstrating bacterial genera and species abundant in gastric cancer (red) compared to healthy controls.

Oesophageal cancer No significant differences in alpha or beta diversity were observed between bacteria from oesophageal cancer tissue, adjacent normal oesophagus and normal oesophagus from healthy control patients (Table 9). Significantly lower concentrations of *Burkholderiaceae* ($p=0.01$) and *Burkholderia*, *Caballeronia* and *Paraburkholderia* ($p=0.02$) were seen in oesophageal cancer bacteria and its adjacent normal oesophagus bacteria, compared to controls (Figure 8). Bacteria classified at a higher taxonomy level such as *Oceanospirillales*, *Oscillospirales*, *Hamomonadaceae* and *Ruminococcaceae* ($p<0.05$) were seen in lower abundances in oesophageal cancer tissue compared to oesophagus from healthy controls. Species such as *Capnocytophaga sputigena* and *Campylobacter gracilis* were more abundant in oesophageal cancer and adjacent normal oesophagus. *Capnocytophaga* are known to harbour the mammalian oropharyngeal tract and contribute to oral disease. Species including *Capnocytophaga sputigena* and *Capnocytophaga leadbetteri* were identified to be significantly higher in samples obtained from oesophageal cancer patients. A detailed list of bacteria identified are listed in Table 10. LEfSe analysis demonstrated a significantly higher abundance of *Howardella*, and 4 species including *Capnocytophaga leadbetteri* and *Prevotella baroniae* (Figure 9).

Table 9. Alpha and beta diversities of bacterial communities in oesophageal cancer tissue, adjacent normal oesophagus, and normal oesophagus from healthy controls. NO = normal oesophagus. P value < 0.05 is considered statistically significant.

	Alpha diversity					Beta diversity				
	Chao 1	ACE	Shannon	Simpson	Fisher	Bray-Curtis	Jensen-Shannon	Jaccard Index	Unweighted Unifrac	Weighted Unifrac
Oesophageal cancer vs Control NO										
Phylum	0.76	0.15	0.79	0.72	-	0.63	1.00	0.34	0.84	0.40
Class	0.80	0.25	0.88	0.85	0.72	0.36	0.98	0.24	0.95	0.41
Order	0.55	0.13	0.82	0.92	0.63	0.38	0.64	0.32	0.49	0.31
Family	0.55	0.18	0.76	1.00	0.49	0.19	0.22	0.16	0.26	0.30
Genus	0.60	0.16	0.76	0.98	0.52	0.25	0.22	0.22	0.29	0.34
Species	0.87	0.32	0.92	0.95	0.82	0.10	0.11	0.09	0.21	0.33
Oesophageal cancer vs Oesophageal cancer NO										
Phylum	0.97	0.48	0.46	0.31	-	0.76	0.95	0.70	0.58	0.64
Class	0.87	0.48	0.54	0.31	0.46	0.86	0.88	0.82	0.59	0.74
Order	1.00	0.87	0.28	0.31	0.42	0.89	0.79	0.96	0.65	0.83
Family	0.95	0.84	0.38	0.38	0.72	0.91	0.86	0.94	0.78	0.87
Genus	0.84	0.62	0.38	0.35	0.92	0.94	0.91	0.96	0.87	0.91
Species	0.82	0.59	0.35	0.31	0.97	0.98	0.97	0.98	0.85	0.91
Oesophageal cancer NO vs Control NO										
Phylum	0.98	0.35	0.90	0.90	-	0.92	0.99	0.67	0.61	0.97
Class	0.68	0.09	0.87	0.99	0.23	0.67	0.88	0.41	0.52	0.89
Order	0.76	0.44	0.36	0.70	0.14	0.53	0.73	0.35	0.24	0.58
Family	0.69	0.26	0.49	0.73	0.20	0.33	0.55	0.24	0.20	0.57
Genus	0.82	0.32	0.64	0.75	0.47	0.31	0.32	0.22	0.18	0.60
Species	0.93	0.42	0.62	0.75	0.67	0.22	0.24	0.18	0.17	0.58

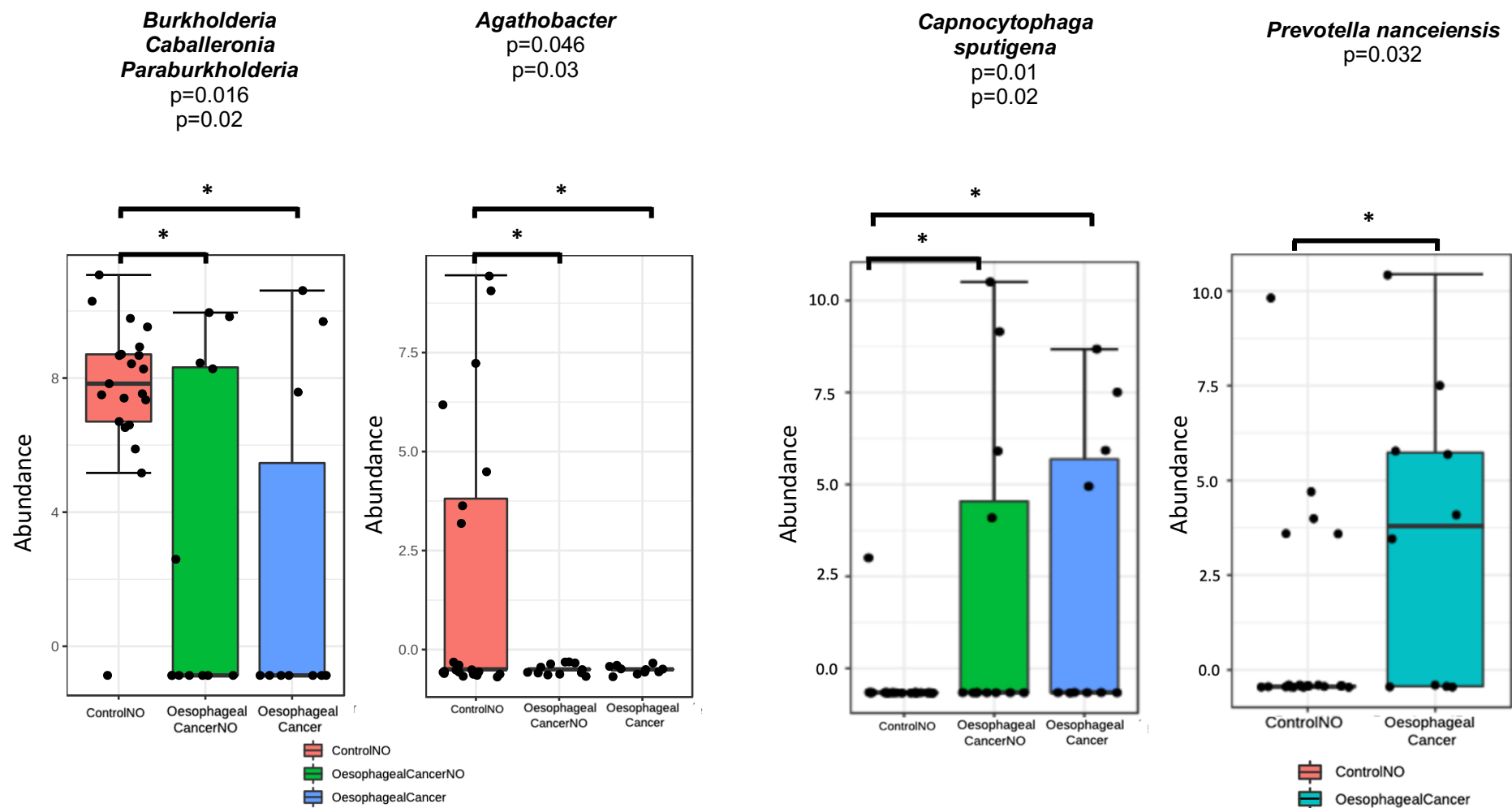


Figure 8. Box plots demonstrating select bacteria enriched in oesophageal cancer tissue compared to healthy control normal oesophagus. P value < 0.05 is considered statistically significant.

Table 10. Bacteria significantly altered in oesophageal cancer-specific bacteria. ↑ = high abundance. ↓ = low abundance.

	P value	In cancer-related bacteria
Oesophageal cancer vs Control NO		
Oceanospirillales (Order)	0.03	↓
Oscillospirales (Order)	0.046	↓
Halomonadaceae (Family)	0.03	↓
Ruminococcaceae (Family)	0.047	↓
Burkholderiaceae (Family)	0.01	↓
Dialister (Genus)	0.02	↑
Halomonas (Genus)	0.03	↓
Howardella (Genus)	0.04	↑
Burkholderia Caballeronia Paraburkholderia (Genus)	0.02	↓
Agathobacter (Genus)	0.046	↓
Lachnoanaerobaculum cf (Species)	0.04	↑
Campylobacter gracilis (Species)	0.04	↑
Capnocytophaga sputigena (Species)	0.01	↑
Capnocytophaga leadbetteri (Species)	0.01	↑
Dialister invisus (Species)	0.02	↑
Prevotella nanceiensis (Species)	0.03	↑
Oesophageal cancer NO vs Control NO		
Thermicanales (Order)	0.04	↑
Aerococcaceae (Family)	0.03	↑
Burkholderiaceae (Family)	0.01	↓
Bacteroidaceae (Family)	0.048	↓
Thermicanaceae (Family)	0.04	↑
Bacteroides (Genus)	0.048	↓
Centipeda (Genus)	0.01	↑
Abiotrophia (Genus)	0.03	↑
Anoxybacillus	0.01	↑
Burkholderia Caballeronia Paraburkholderia (Genus)	0.02	↓
Faecalibacterium	0.048	↓
Thermicanus	0.04	↑
Lachnoanaerobaculum	0.03	↑
Mogibacterium	0.03	↑
Agathobacter (Genus)	0.03	↓
Lachnoanaerobaculum cf (Species)	0.006	↑
Campylobacter gracilis (Species)	0.02	↑
Capnocytophaga sputigena (Species)	0.02	↑
Gemella parahaemolysans (Species)	0.008	↑
Abiotrophia defectiva (Species)	0.03	↑
Bifidobacterium dentium (Species)	0.04	↑
Faecalibacterium prausnitzii (Species)	0.048	↓

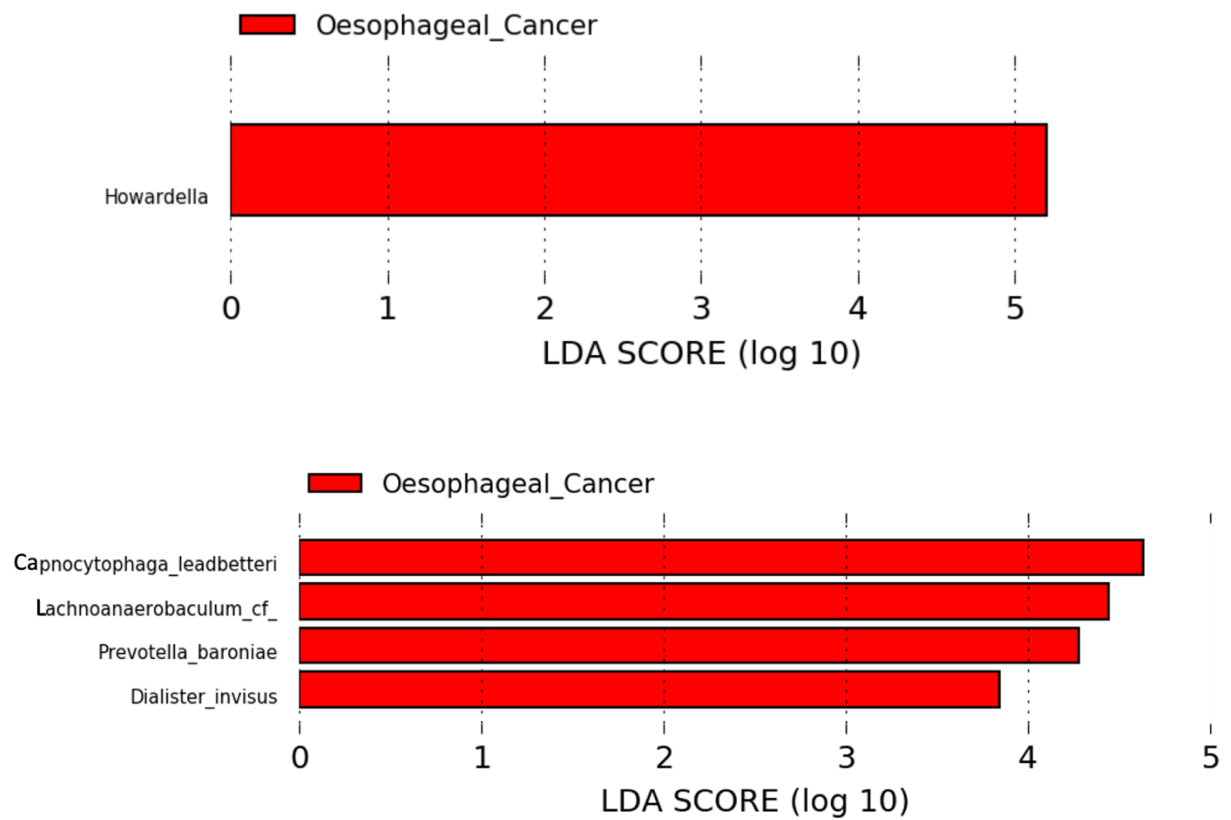


Figure 9. Linear discriminant analysis Effect Size demonstrating bacterial genera and species abundant in oesophageal cancer (red) compared to healthy controls.

4.5 DISCUSSION

This study primarily identified (i) bacteria related to cancer tissue in comparison to healthy controls and (ii) bacteria related to tissue adjacent to cancer in comparison to healthy controls. Secondary findings were of bacteria linked to contamination from commercial DNA extraction kits, particularly pronounced in low biomass samples.

Gastric adenocarcinoma samples yielded higher abundances of *Lactobacillus* at all taxonomic levels, including that of tissue adjacent to the tumour. These findings are consistent with literature, highlighted in the review in Chapter 2. Fourteen independent studies reported higher abundances of *Lactobacillus* in gastric cancer, spanning various matrices including tissue and saliva. This is important in validating the quality of the methodology used. *Lactobacillus* primarily thrives in microaerophilic environments which is unique to the oesophagogastric tract. *Lactobacillus* is a recognised probiotic globally, which contradicts these findings. However, the functional capabilities of bacteria in pathogenic settings may be altered. This is currently unknown. *Pseudopropionibacterium*, a Gram-positive aerotolerant anaerobe, was also identified in this dataset. This is potentially important as *Pseudopropionibacterium* is strongly linked to the production of acetic acid, propionic acid and lactic acid following glucose stimulation. This is well-established in the food, cosmetic and pharmaceutical industries as detailed in Chapter 3. This review highlights that both *Lactobacillus* and *Propionibacterium* can generate a higher yield of acetic acid and propionic acid on an industrial scale when exposed to glucose as a stimulant. In the context of this work, short chain fatty acids (C2-C6) have been detected in the exhaled breath of patients with oesophagogastric cancer. This supports a possible hypothesis that the origin of the short chain fatty acids may be cancer-specific bacteria. Other bacteria demonstrated in this study consistent with the literature include *Bacilli* and *Lactobacillaceae*. Six *Lactobacillus* species were highlighted in the review (*L. amylolyticus*, *L. coleohominis*, *L. fermentum*, *L. gasseri*, *L. reuteri*, *L. salivarius*), however this study has identified a new species in gastric cancer; *L. vaginalis*.

In oesophageal cancer, *Prevotella* species were found in higher abundances in cancer tissue compared to healthy controls, in keeping with the findings in the literature. Remaining analysis of oesophageal cancer related bacteria demonstrated novel bacteria not reported in literature previously. *Capnocytophaga* species such as *Capnocytophaga sputigena* and *Capnocytophaga leadbetteri* were highly abundant in oesophageal cancer tissue but have not been reported in literature previously. Many of these bacteria were found in lower abundances in oesophageal cancer patients compared to healthy controls. It is uncertain whether these bacteria are downregulated in the cancer population or upregulated in healthy controls.

The main limitation of this study is the introduction of contaminants by use of commercially available DNA extraction kits. Although this does not present a problem for certain samples such as faeces which are high bacterial biomass, for low biomass samples from the upper gastrointestinal tract, contaminants are amplified. All bacteria identified were screened and previously reported contaminants (and through

experience) were excluded, such as *Delftia*. New, less common bacteria such as *Howardella*, *Oscillospirales* and *BCP*, need to be investigated and interpreted with caution. Confounding factors in microbiome studies such as chemotherapy and proton pump inhibitor use were acknowledged, but not used as an exclusion criterion in this discovery study. For future studies this will need to be considered.

4.6 CONCLUSION

The microbial community identified in this cohort of patients with oesophagogastric cancer holds some similarity to those described in literature. *Lactobacillus* and *Prevotella* were among the top five found to be highly abundant in gastric cancer, at more than one taxonomic level. *Pseudopropionibacterium* in this study, and *Propionibacterium* in literature were also prevalent. There were fewer trends observed within the oesophageal cancer group and this study adds more novel bacteria, such as *Capnocytophaga species*, not previously reported. In normal practice, the oesophagus and stomach are grouped together, and the microbial ecosystem is considered to be shared, however; findings from this study and literature demonstrate distinct differences in the bacterial environments.

5. CHAPTER 5: BACTERIA CELL CULTURE FROM OESOPHAGOGASTRIC CANCER TISSUE AND SALIVA

5.1 BACTERIA CELL CULTURES FROM OESOPHAGOGASTRIC CANCER TISSUE AND SALIVA USING SHOTGUN METAGENOMIC ANALYSIS

5.1.1 Introduction

Historically, bacterial cell culturing methods were the mainstay of identifying bacterial species. This presented a challenge for bacteria requiring strict anaerobic culture conditions or co-dependency of other bacterium types. However, recent advancements of culture independent methods such as 16S rRNA gene sequencing and metagenomic workflows have revolutionised the characterisation of gastrointestinal microbiota. Bacteria cell culturing still plays a pivotal role in biotransformation processes, and in our interest, in exploiting the volatile organic compounds emitted from cancer specific bacteria. Bacteria cultured in optimal physiological conditions can provide an insight into the end metabolite profiles which may be crucial in identifying volatile biomarkers. *In vitro* studies are an important adjunct in understanding *in vivo* experimental outcomes.

5.1.2 Aims

The primary aim of this study is to establish mixed and pure bacteria cell cultures from oesophagogastric cancer tissue and saliva with identification through metagenomic whole genome sequencing.

Secondary aims included:

- (i) Exploration of the metabolic and functional capabilities of cancer-specific bacteria.
- (ii) Investigation of a link between the oral and cancer tissue microbiome.

5.1.3 Methods

Patient selection and Sample preparation, collection and storage are detailed in Section 2, Chapter 4.

Bacteria cell culture from Tissue: Glycerol stocks containing tissue samples were thawed within a biological safety class II hood. Class II containment procedures were followed for culturing unidentified organisms. Tissue was aseptically removed from the glycerol stock using pipette suction and placed in a 1.5mL screw-cap tube with 300µL of sterile PBS pH 5.0. PBS pH 5.0 was selected to represent the gastric environment in a patient medicated with a proton pump inhibitor, which embodies much of this patient cohort. Each sample was homogenised by vortexing at a maximum speed of 1800 rpm for 30 seconds, inverting 3 times, and further vortexing for 30 seconds. This step gently removes the mucosal bacteria from the tissue. The tissue was removed by pipette suction. 150µL of this suspension was

transferred onto a fastidious anaerobic agar (FAA) plate and distributed evenly using a L-spreader. The remaining 150µL was preserved for 16S rRNA gene sequencing to assess culture methods and the representativeness of the bacteria cultured. FAA plates were sealed and inserted into an EZ anaerobic pouch and incubated at 37°C for 4 days. The anaerobic pouch was changed at 48 hours.

Pure cultures: Bacterial colonies were morphologically characterised and up to 4 predominant colonies were picked with a 1µL loop, resuspended in 150µL of sterile PBS pH 5.0 and spread onto a fresh FAA plate to establish pure cultures. The same steps for cultures were followed as previously described.

Mixed cultures: The remaining colonies were scraped off the agar and pooled into a mixture of sterile fastidious anaerobic broth/7% horse blood. All samples were frozen at -80°C.

Bacterial cell culture from Saliva: Glycerol stocks were thawed and homogenised by vortexing for 30 seconds, inverted 3 times, and vortexed for a further 30 seconds at 1800 rpm. 150µL was spread onto a nutrient agar plate to encourage growth of aerobic microorganisms, and a further 150µL onto a FAA plate for anaerobic growth. The sealed plates were incubated at 37°C for 48 hours. Anaerobic cultures produced heavy growth making it difficult to identify individual colonies. Serial dilutions from 10^{-4} to 10^{-7} (50µL in 450µL sterile PBS) were plated and growth assessed at 24 and 48 hours. Slower growing colonies were observed at 48 hours. No macroscopic difference was seen with a further 2 days of incubation, totalling 4 days (Figures 10 and 11). It was decided additional incubation time may limit nutrients available for growth, and therefore an incubation periods of 48 hours were selected. Adequate separation of colonies was reported at 10^{-5} dilution. Considering variability between sample types, 10^{-4} dilutions (1:10,000) were used for anaerobic culture. Aerobic cultures showed sparse growth with dilution, and therefore the study proceeded with undiluted mixtures. The remaining saliva-glycerol stock mixture was preserved for 16S rRNA gene sequencing. Pure and mixed cultures were obtained as per protocol detailed above. Saliva cultures were maintained in PBS stock.

Cell bank characterisation The integrity of bacterial glycerol stocks were assessed. Two mixed populations were thawed, plated onto FAA plates, and incubated as per protocol for 4 days at 37°C. Morphology was compared with the original sample growth. Macroscopic assessment suggested the chosen media for cell banking was successful.

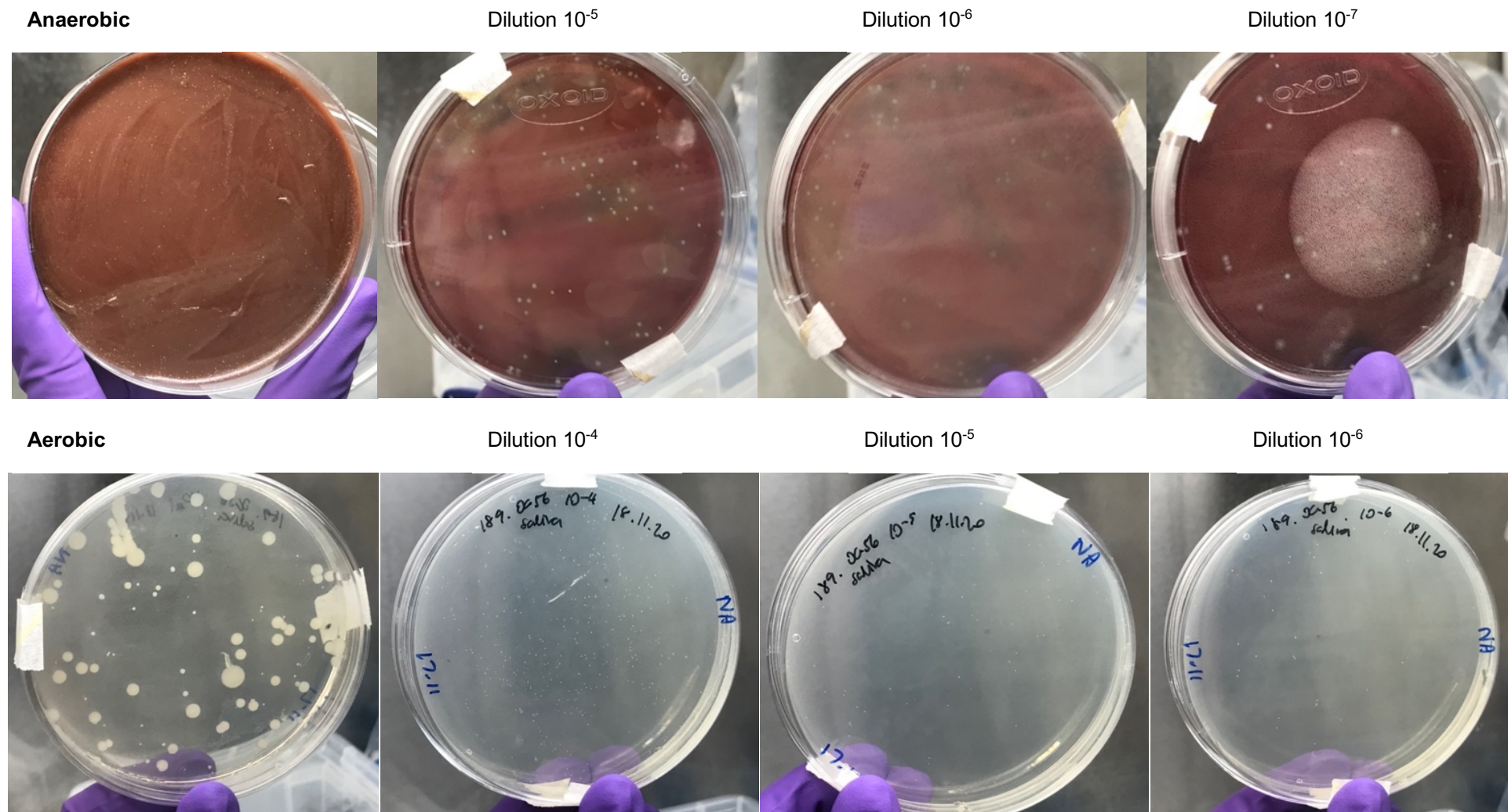
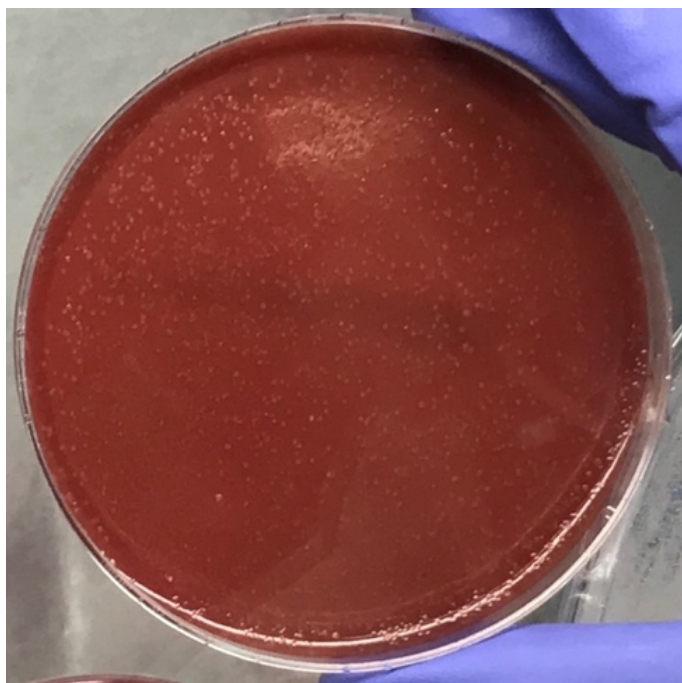


Figure 10. Anaerobic and aerobic saliva growth with serial dilutions from 10^{-4} to 10^{-7} (50 μ L in 450 μ L sterile PBS) to separate colonies on fastidious anaerobic agar and nutrient agar, respectively. Sample 189 – OG56. Slower growing colonies were observed at 48 hours. No macroscopic difference was seen with a further 2 days of incubation, totalling 4 days.

Original culture
30 October 2020



Bacteria-glycerol stock
20 November 2020

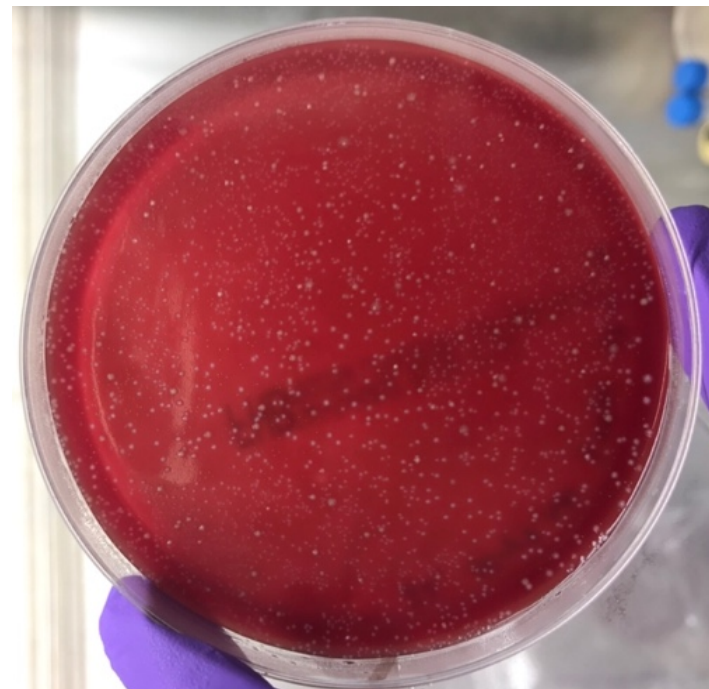


Figure 11. Macroscopic assessment of bacteria-glycerol stocks demonstrating a similar growth pattern to the original culture, following growth on fastidious anaerobic agar. Sample 33 OG47 NO.

Bacteria DNA extraction from cultures Selected mixed and pure bacteria-glycerol stocks were thawed and centrifuged at 10,000 g for 10 minutes. The supernatant was discarded, and the bacteria cell pellet resuspended in PBS up to 200µL. The FastDNA Spin® (MP Biomedicals) kit was used for DNA extraction. The bacteria suspension was transferred to a Lysing Matrix A tube with 1mL cell lysis solution for tissue and cells (CLS-TC) and homogenised in a Next Advance Bullet Blender® Storm 24 bead beater for 40 second at a speed 6.0. The samples were centrifuged at 14,000 g for 10 minutes to pellet the debris. 800µL of the supernatant was added to an equal volume of the binding matrix, inverted to mix, and incubated with gentle agitation for 5 minutes at room temperature. 400µL was centrifuged twice through the SPIN filter at 14,000 g for 1 minute. The pellet was resuspended in 500µL of SEWS-M prepared with 100% ethanol. The suspension was further centrifuged at 14,000 g for 1 minute, contents discarded, and centrifuged for a further 2 minutes. 100µL of DES was used to elute the DNA by resuspending the binding matrix and incubating the sample for 5 minutes at 55°C. After a final centrifugation at 14,000g for 1 minute, the DNA was ready for downstream application. Samples were stored at -80°C until further processing.

Metagenomic Sequencing Mixed bacteria culture samples underwent Shotgun metagenomic sequencing using the Illumina Nextera XT kit and Illumina Novaseq 150bp PE (2Gb data per sample). Pure cultures underwent Illumina 16S rRNA gene (V3-V4) amplification and Illumina Miseq sequencing 300 PE (approximately 100,000 reads per sample; Macrogen®, South Korea).

Characterisation of metagenomes derived from mixed bacterial cultivation work. The quality of the raw sequence data (paired-end 150 bp) was assessed using fastQC v0.11.9; the data did not require trimming. An average of 2.91 Gb ($\pm$ 0.39 Gb) of sequence data were generated for each sample. To determine taxonomic coverage within samples, paired-end read data were run through Kraken2 2.0.8-beta [186] using the pre-compiled Kraken2 GTBD_r89_54k index (downloaded on 3 May 2020) available from <https://bridges.monash.edu/ndownloader/files/16378439> [187], with data summarized at different taxonomic levels (phylum to species). This gives a broad overview of microbes present within a sample. To generate sample-specific data, metagenome-assembled genomes (MAGs) needed to be assembled. The sequence data were assembled using MEGAHIT v1.2.9 [188]. Raw sequence reads were mapped against the assembled contigs and MetaBAT2 v2.12.1 was used to bin the data [189]. A total of 727 bins were created across all samples. MAGpurify v2.1.2 (database v1.0) was used to clean up the bins (i.e. the MAGs) [190]. The taxonomic affiliations of the microbes represented in the MAGs were determined using ANI analysis with GTDB-Tk v1.3.0 (release 95).[191] Assessments of completeness and contamination of the MAGs were made using CheckM v1.13.0; MAGs were annotated using Prokka v1.14.0 and summary statistics were determined for all. Overall quality of MAGs was assessed according to Bowers *et al.*: those with ≥ 90 % completeness and ≤ 5 % contamination were considered high-quality draft genomes; those with ≥ 50 % completeness and ≤ 10 % contamination were considered medium-quality draft genomes; all other MAGs were considered low quality and were not examined further.[192] A phylogenetic tree incorporating the medium- and high-quality MAGs was generated using PhyloPhlAn v0.99 and visualized using iTOL v6.3.[193] The primary metabolites

encoded within the medium- and high-quality MAGs were predicted using gutSMASH (Pfam_A.hmm, 2 May 2020).[194]

Characterization of pure isolates recovered from bacteria cultivation work. Trimmed raw sequence data (microbesNG) were checked using fastQC v0.11.9. Further editing was not required. Sequence data were assembled using SPAdes v3.11.1.[195] Contigs <200 nucleotides in length were removed from assemblies using reformat.sh in BBMap. Prokka v1.14.0 was used to predict coding sequences encoded within the draft genomes.[196] CheckM v1.1.3 was used to determine the completeness and contamination of the genomes.[197] Tentative identities for the assembled genomes were determined using PubMLST Species ID.[198] Taxonomy and metadata files for release 202 of the Genome Taxonomy Database were downloaded, and used to identify type species sequences for taxa most closely related to the three isolates sequenced by microbesNG.[199] The relevant 113 representative high-quality genomes were downloaded from GenBank and used to create a database against which the three newly generated genome sequences were compared: (i) for average nucleotide identity (ANI) analysis using fastANI [200], then (ii) phylogenetic analysis using PhyloPhlAn v0.99.[201]

5.1.4 Results

Taxonomic characterization of microbes detected in metagenomic data recovered from mixed bacterial cultivation work Shotgun metagenomic data were generated for 68 samples. After quality checks, assembly and binning of data, a total of 727 MAGs were created across all samples (Figure 12). Only the medium- and high-quality MAGs were analysed. Taxonomic analysis of the 288 medium- and high-quality MAGs showed they belonged to eight different phyla, with two MAGs not assigned to any taxonomic rank based on the gtdb_tk-based analysis (Figure 13). Sixty-five of the 68 samples processed harboured medium- to high-quality MAGs. Phylogenetic analyses of the 288 MAGs allowed the two unassigned MAGs to be assigned to the phyla *Proteobacteria* and *Firmicutes_A*, respectively (Figure 14). The eight phyla detected in the 65 samples covered 43 genera: predominantly 8 genera *Streptococcus*, *Prevotella*, *Veillonella*, *Rothia*, *Granulicatella*, *Neisseria*, *Pauljensenia* and *Haemophilus* (Figure 15). In total, the medium- and high-quality MAGs recovered from the 65 samples represented 109 different species of bacteria.

Prediction of primary metabolites encoded by the medium- and high-quality MAGs gutSMASH was used to predict the primary metabolites potentially produced by the 288 MAGs, with 274/288 MAGs returning hits (Figure 16).

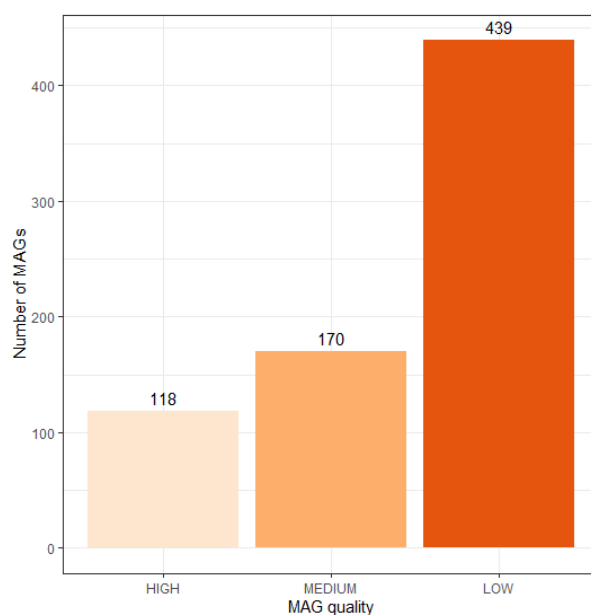


Figure 12. Number and quality of MAGs obtained from metagenomes generated from mixed bacterial cultivation work.

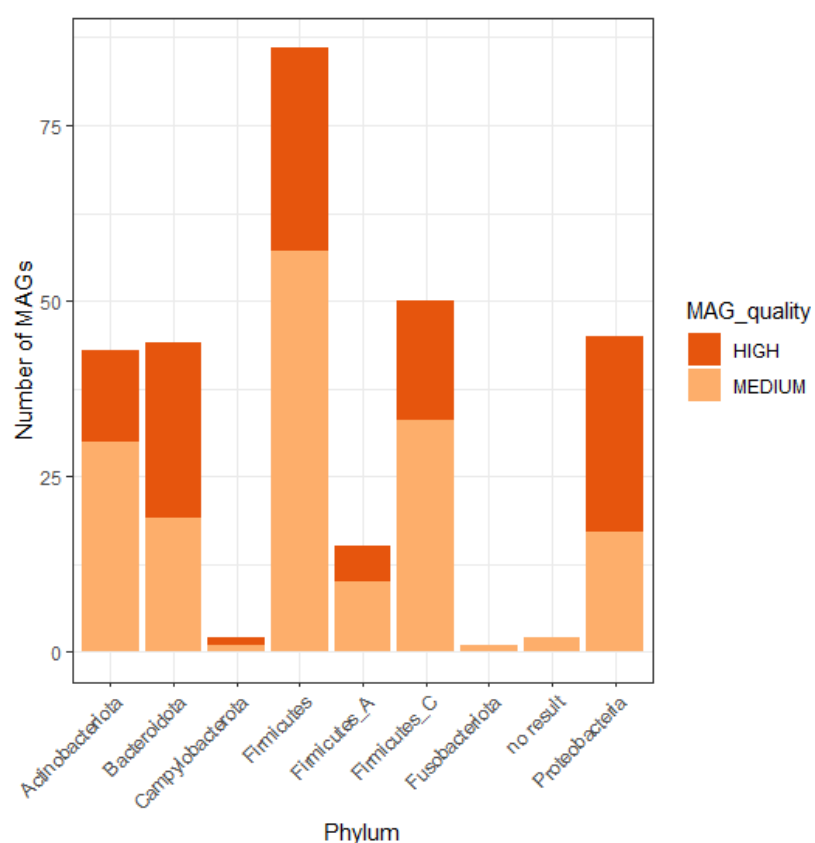


Figure 13. Phylum-level breakdown of the medium and high-quality MAGs generated from original cultivation work. n= 170 medium quality and n=118 high quality according to Bowers *et al.*

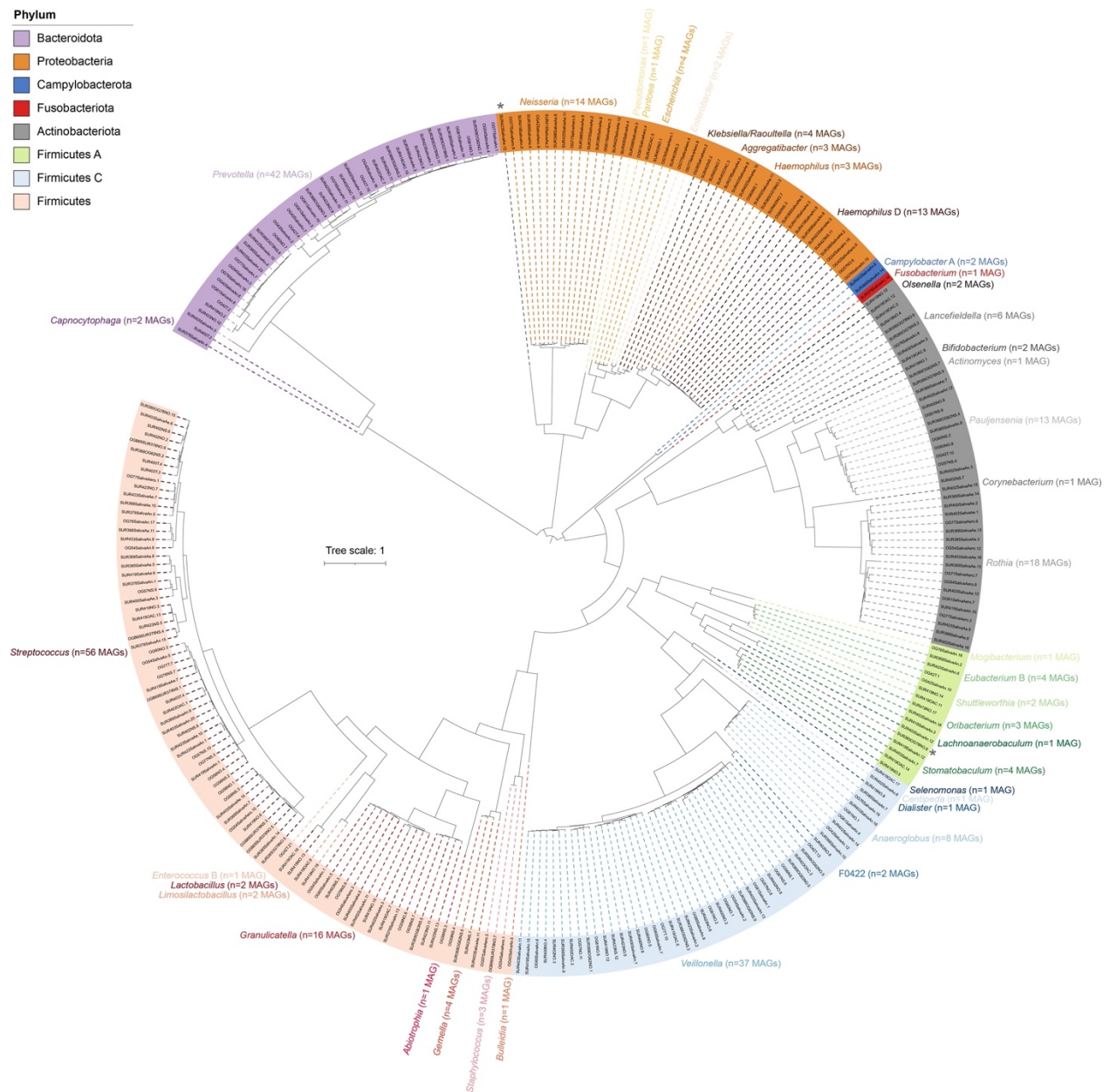
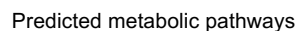


Figure 14. Phylogenetic tree, generated using PhyloPhlAn, showing the taxonomic affiliations of the medium- and high-quality MAGs recovered from 65/68 of the metagenomes derived from the original isolation work. Solid colours (corresponding to the legend) represent the phyla. Genera are labelled around the outside of the phylogenetic tree. Scale bar, normalised fraction of total branch length.



results.

16S gene Sangar Sequencing to confirm pure bacterial cultures Chromatograms were quality assessed visually demonstrating clean single peaks. Ab1 files were opened in 4 peaks software and the start/end primer attachments were deleted. The sequence was entered into NCBI Blast: Nucleotide sequence and EZbiocloud databases for bacterial identification; (a) *Escherichia coli* (b) *Streptococcus salivarius* and *Streptococcus vestibularis* and (c) *Veillonella infantium*.

Characterization of three isolates recovered from cultivation work Sample 46 (46Esc) was predicted to encode 4596 genes (plus 11 rRNAs, 2 repeat regions, 80 tRNAs and 1 tmRNA), sample 49 (49Vei) was predicted to encode 2056 genes (plus 6 rRNAs, 48 tRNAs and 1 tmRNA) and sample 50 (50Str) was predicted to encode 2613 genes (plus 7 rRNAs, 56 tRNAs and 1 tmRNA). All three genomes were of high quality (in terms of completeness and contamination, and the criteria of Bowers *et al.* [192]), though the genome of 50Str was highly fragmented (based on number of contigs) (Table 11, Figure 17).

Table 11. Summary statistics for the draft genome sequences generated in this study.

Strain	Tentative ID*	No. contigs	Genome size (nt)	N50	Completeness (%)	Contamination (%)
46Esc	<i>Escherichia coli</i> 100 %	72	4951507	550503	99.97	0.33
49Vei	<i>Veillonella dispar</i> 89 %	170	2274247	181949	100.00	2.99
50Str	<i>Streptococcus salivarius</i> 98 %	1844	3412628	114096	99.96	2.86

*Determined using PubMLST Species ID.

PubMLST (public databases or molecular typing and microbial genome diversity) had provided tentative identity information for the isolates but cannot be trusted to provide absolute identification of sequence data, especially for genomes in many contigs (such as that of 50Str). Consequently, ANI values were determined for the three genomes against relevant related taxa. 46Esc shared 98.64 % ANI with *Escherichia coli* ATCC 11775^T; 49Vei shared 94.58 % ANI with *Veillonella dispar* NCTC 11831^T; 50Str shared 95.88 % ANI with *Streptococcus salivarius* NCTC 8618^T. ANI values of ≥95 % between two genomes are indicative of them representing the same species, with affiliations confirmed via phylogenetic analyses.[202-204] On the basis of the ANI data presented here, 46Esc is a strain of *E. coli* and 50Str is a strain of *S. salivarius*, while 49Vei represents a novel species of the genus *Veillonella*. Phylogenetic analysis (Figure 18) confirmed these findings.

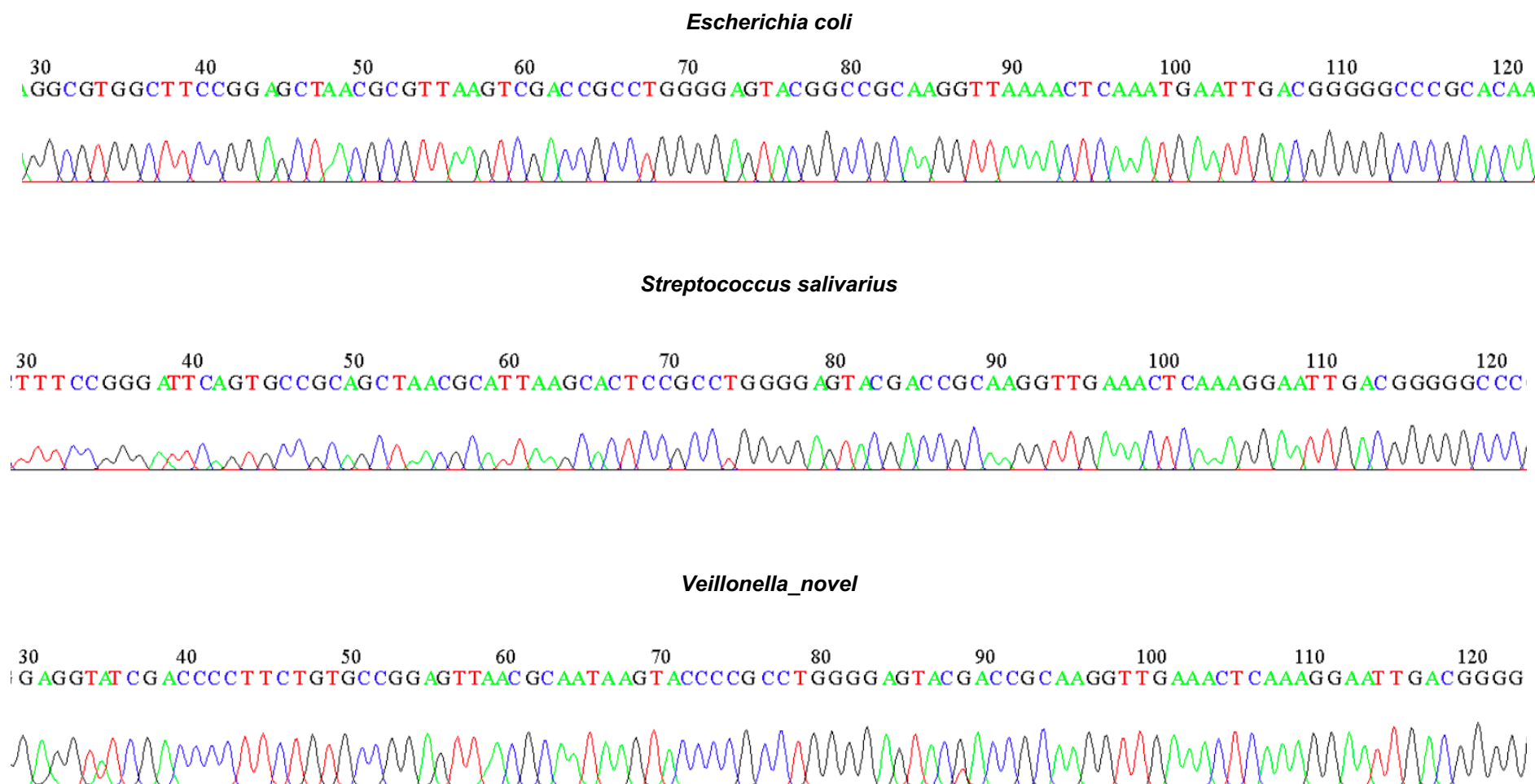


Figure 17. Chromatograms demonstrating good quality peaks representative of pure cultures. All three bacteria cultures were established directly from gastroesophageal junctional cancer tissue.

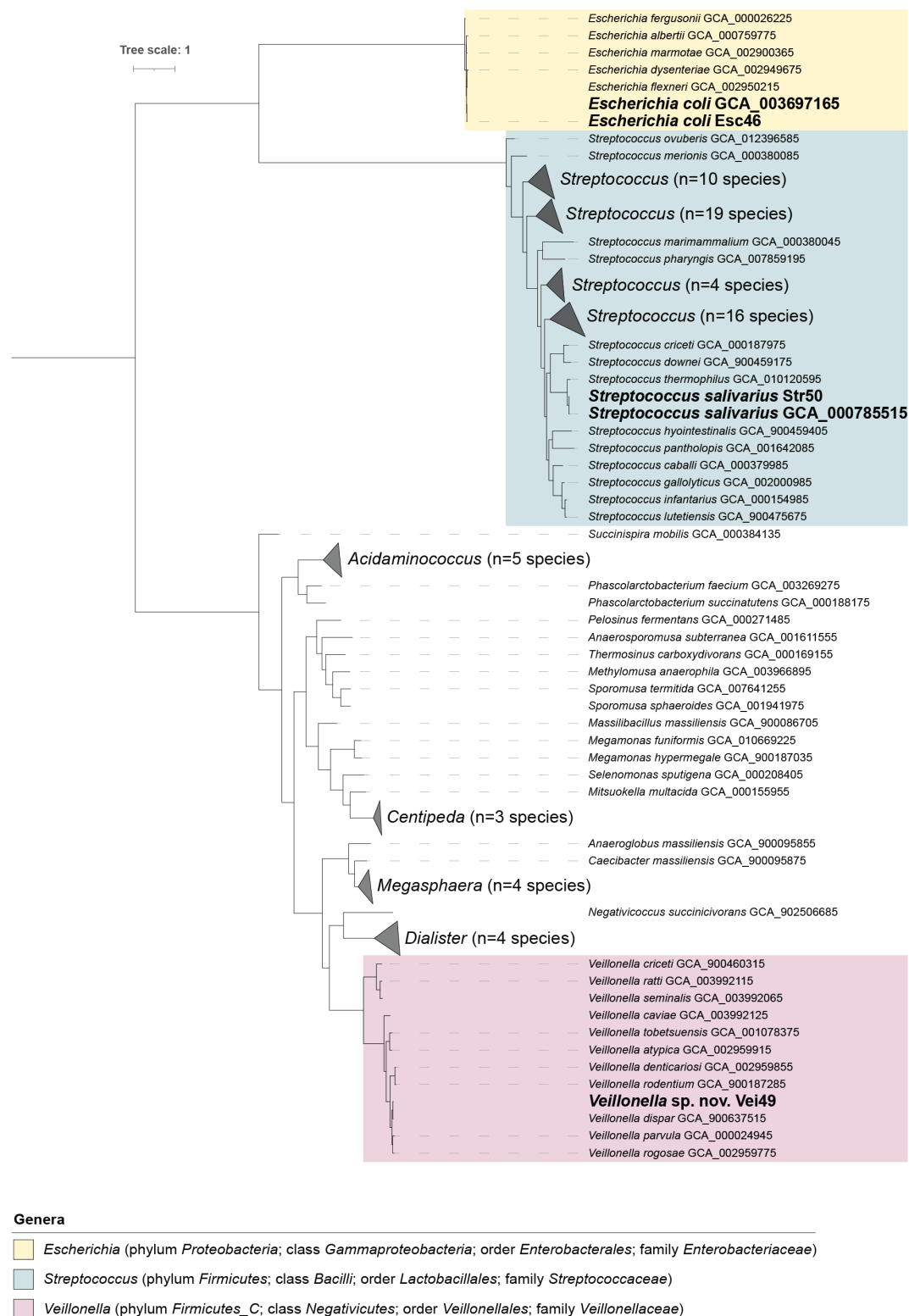


Figure 18. Phylogenetic tree showing the relationships of the three isolates' genomes with 113 related taxa. Clades of some multi-species genera have been collapsed to aid visualisation. The tree was generated using PhyloPhlAn v0.99 and visualized using iTOL. Scale bar, normalised fraction of total branch length.

Association between the tissue and saliva microbiome

The microbial distribution of the upper gastrointestinal tract tissues and saliva was explored to assess for potential correlative relationships which may provide a platform for (i) non-invasive diagnostics using the saliva microbiome and (ii) explain the contribution of the oral microbiome in the augmentation of target VOCs. Current literature suggests that the oral microbiome is altered in cancer and can therefore be used for diagnostic purposes. However, to the best of our knowledge no study has explored the association between the saliva and upper gastrointestinal tissue microbiome. This study assessed the bacterial correlation at the genus and the species level in three patient groups: gastric cancer, oesophageal cancer and healthy controls.

Association by Genera 3 genera were dominant in the saliva and cancer tissue of gastric cancer patients (n=3): *Eubacterium*, *Prevotella*, *Streptococcus*. In oesophageal cancer (n=3), 4 genera were identified: *Prevotella*, *Veillonella*, *Streptococcus*, *Stomatobaculum*. In the healthy control group (n=5), 4 genera were also shown to be present in both saliva and tissue: *Prevotella*, *Streptococcus*, *Veillonella*, *Anaeroglobus*.

Association by Species Five bacterial species were identified in saliva and associated gastric cancer tissues: *Streptococcus parasanguinis*, *Prevotella pallens*, *Prevotella salivae*, *Prevotella jejunii*, *Eubacterium sulci*. Two bacterial species were found in the oesophageal cancer group: *Veillonella atypica*, *Stomatobaculum longus*. Five species were present in the control group (n=4): *Veillonella atypica*, *Veillonella dispar*, *Prevotella melaninogenica*, *Anaeroglobus micronuciformis*, *Streptococcus contellatus*.

Predicted primary metabolic pathways

gutSMASH has identified several potential metabolic pathways within the mixed and pure bacterial cultures. The similarity by percentage identity to known gene clusters within the database was provided. The 10 most common pathways present were: pyruvate to acetate/formate (100%), fumarate to succinate (100%), succinate to propionate (100%), acetate to butyrate (100%), glutamate to butyric (90%), threonine to propionate (70%), PFOR II (pyruvate ferredoxin oxidoreductase; 60%), porA (60%), ethanolamine utilisation (EUT; 47%), propanediol utilisation (PDU; 50%).

Pyruvate to acetate and formate pathway The pyruvate to acetate/formate pathway is well established in bacterial metabolism. Glucose is catabolised by NAD⁺ enzyme to produce pyruvate. Pyruvate undergoes two distinct downward metabolic processes (i) decarboxylation by enzyme CoA to produce acetyl-CoA and subsequently end metabolite acetate, (ii) pyruvate formate lyase (PFL; glycyl-radical enzyme) converts pyruvate to formate in anaerobic conditions. During fermentation, equal amounts of acetate and ethanol are produced from acetyl-CoA. Phosphotransacetylase (PTA) and acetate kinase (ACK) converts acetyl-CoA into acetate. Acetyl-CoA is reduced by aldehyde

dehydrogenase (ALDH) and alcohol dehydrogenase (ADH) to produce ethanol. A minor by product of this pathway is lactate, produced by the catalytic conversion of pyruvate by the lactate dehydrogenase (LDH) enzyme. Overall, there are four end metabolic products produced from the activation of this pathway in anaerobic conditions: acetate, ethanol, formate and lactate. This pathway demonstrated a 100% identity match to known gene clusters.

Fumarate to succinate pathway The fumarate to succinate pathway dominated many of the samples with 100% identity to known gene clusters, with a known end metabolite propionate. Following glycolysis, 2-phosphoenolpyruvate (2-PEP) is converted to malate by malate dehydrogenase, and subsequently to fumarate by fumarase (Fum). The final step of the pathway uses fumarate reductase (Frd) to convert fumarate to succinate. This process occurs anaerobically, and the activity of Fum is halted in the presence of oxygen.

Succinate to propionate pathway The succinate to propionate pathway has been identified in the majority of bacteria samples and is recognised to be an important well established pathway in the production of propionic acid via bacterial metabolism. This gene was a 100% match to known gene clusters. Carbon energy sources, such as monosaccharides, are metabolised to PEP and subsequently oxidised to oxaloacetate. The conversion to malate, fumarate and succinate is described above in the fumarate to succinate pathway. Further degradation by the enzyme methyl malonyl-CoA decarboxylase subunit alpha (mmdA) leads to the production of propionate. An alternative two pathways exist: (i) lactate to propionate production and (ii) propanediol pathway.

Acetate to butyrate pathway Butyrate is a predominant metabolite in bacterial metabolism. Carbohydrate catabolism produces pyruvate followed by acetyl-CoA after decarboxylation by CoA enzyme. Two molecules of acetyl-CoA combine to form butyryl-CoA. Phosphotransbutyrylase converts butyryl-CoA to end metabolite butyrate. This is a well-established pathway with a 100% identity to known gene clusters.

Glutamate to butyric pathway Butyric acid is a common metabolite in the gastrointestinal tract with potentially differing activities: essential for the healthy gut but toxic in the oral cavity associated with periodontal disease and oral cancer.[205] This pathway was identified with 90% similarity to known gene clusters. Glutamate is catabolised to 2-oxoglutarate by the enzyme glutamate dehydrogenase (Gdh) with ammonia by-product. L-2-hydroxyglutarate dehydrogenase (L2Hgdh) further reduces 2-oxoglutarate to 2-hydroxyglutarate, and subsequently to 2-hydroxyglutaryl-CoA by glutaconate CoA-transferase (Gct). The following sequence of reactions take place: 2-hydroxy-glutarylCoA dehydrogenase (HgCoAd) enzyme metabolism produces glutaconyl-CoA; glutaconyl-CoA decarboxylase (Gcd) produces crotonyl-CoA; butyryl-CoA dehydrogenase (Bcd) enzyme produces butyryl-CoA and finally butyrate.

Threonine to propionate pathway The threonine to propionate pathway is active in anaerobic conditions. L-threonine is converted into 2-ketobutyrate by L-threonine dehydrate catabolic B (TdcB)

enzyme. TdcE deaminates 2-ketobutyrate to propionyl-CoA and formate. Phosphate acetyltransferase (Pta) enzyme converts propionyl-CoA into propionyl-P and TdcD facilitates the final catalysis into propionate. Propionyl-CoA can be converted into acetate by acrylyl-CoA enzyme. The end metabolic products are propionate and acetate. This pathway has a 70% match to known gene clusters.

Pyruvate ferredoxin oxidoreductase (PFOR II) pathway This pathway is present in anaerobic metabolism. The decarboxylation of pyruvate is catalysed by PFOR enzyme to produce acetyl-CoA. Downstream into the TCA cycle produces propionate and butyrate. This pathway has a 60% identity match to known gene clusters.

porA pathway This pathway uses a long-chain polypeptide to be catabolised in a process to produce short chain fatty acids. This pathway has a 60% similarity to known gene clusters.

Ethanolamine utilisation (EUT) pathway The EUT pathway utilises ethanolamine which is present within the cell membranes lining the gastrointestinal tract. Ethanolamine is a carbon source for bacteria and is catabolised by EutH enzyme into acetaldehyde. Further downstream processes include: (i) acetaldehyde conversion into acetyl-CoA by EutE, to acetyl-P by EutD, into acetate by acetate kinase (Ack) enzyme; (ii) acetyl-CoA into the TCA cycle to produce butyrate; (iii) acetaldehyde conversion into ethanol by EutG. There is a known 47% match to known gene clusters.

Propanediol utilisation (PDU) pathway The propanediol pathway has a known 50% similarity to known gene clusters. The end metabolite is propionate. A carbon source such as glycerol is converted into 3-hydroxypropionaldehyde (3-HPA). Various Pdu enzymes are involved in the conversion to 3-hydroxypropionyl-CoA, 3-hydroxypropionyl phosphate, 3-hydroxypropionate, 1,2-propanediol and propanal. Two further downstream processes include: (i) conversion to propanol, (ii) conversion to propionyl-CoA, propionyl phosphate, and finally propionate.

5.1.5 Discussion

The primary findings of this study are: (i) high abundances of *Streptococcus*, *Prevotella*, *Veillonella*, *Rothia*, *Granulicatella*, *Neisseria*, *Pauljensia* and *Haemophilus* in the upper gastrointestinal tract; (ii) the association between saliva and gastric cancer tissue: *Streptococcus parasanguinis*, *Prevotella pallens*, *Prevotella salivae*, *Prevotella jejunii* and *Eubacterium sulci*; (iii) the association between saliva and oesophageal cancer tissue: *Veillonella atypica* and *Stomatobaculum longus*; (iv) identification of potential metabolic pathways and their predicted metabolites.

The diagnostic potential of the saliva microbiome and the cancer tissue microbiome have been explored separately with promising results. Research demonstrates that the oral microbiome is reflective of cancer, in gastric, colorectal and pancreatic cancers, with a potential for a screening tool.[38, 51, 52] However, to date, no direct associations have been made between the saliva and gastrointestinal tract cancers. This study identified a panel of bacterial species found in both saliva and tissue matrices in the same cancer patient. In the gastric cancer group, five species (*Streptococcus parasanguinis*, *Prevotella pallens*, *Prevotella salivae*, *Prevotella jejunii* and *Eubacterium sulci*) were recognised. On the genus level, *Streptococcus* and *Prevotella* have been demonstrated to be highly abundant bacteria in cancer tissue and from the oral cavity (saliva, tongue coating) in literature. However, on a species level, only *Streptococcus parasanguinis* has been isolated from gastric cancer tissue. However, 16S rRNA gene analysis is more commonly used for genus level analysis which may explain why there is sparse species level data available. A similar observation was made in the oesophageal cancer group. *Veillonella* was identified in cancer tissue, however species level data was not comparable. On a genus level, cancer associated bacteria, such as *Veillonella*, *Streptococcus* and *Prevotella*, are in keeping with literature findings. Translocation of bacteria from the mouth may be responsible for similarities in the bacterial profile seen in the oesophagus or stomach. However, the functionality of the bacteria may be different, predominantly affected by the microenvironment where the oral cavity is exposed to an aerobic and alkaline environment. In contrast the oesophagogastric tract is more acidic and considered microaerophilic. Therefore, abundance, functionality and the metabolic capabilities of bacteria need to be further explored with a multi-omics approach to define the cancer-specific bacterial profile.

It is currently unknown whether cancer cells cause changes in the microbiome or alternatively if the microbiome change causes cancer pathogenesis. The change in the luminal microenvironment may influence the bacterial composition in the surrounding tissue adjacent to the cancer and beyond. Defining the relationship between the cancer tissue and oral microbiome may have useful implications for future translation of diagnostic biomarkers into clinical practice.

This study has provided bacterial abundance data; however, this does not reflect their functional capabilities. The predicted pathways supporting the proof of concept that these specific bacteria isolated from oesophagogastric cancer have the potential to produce SCFA, with up to 100% identity match to known gene clusters. The intermediate pyruvate is linked to acetate production via the

pyruvate to acetate/formate pathway. Pyruvate is also known to be pivotal in glycolysis and the production of butyric acid, acetic acid and ethanol. This has been well described in the biofuel industry to produce the highest yields of SCFAs used on an industrial scale. Propionate production is predominantly the end-product of the succinate pathway which is well established in the bacterial community. Other pathways such as acetate to butyrate and glutamate to butyric are also well defined in bacterial metabolism. The role of less common pathways such as PFOR II, porA, EUT and PDU pathways require more investigation.

5.1.6 Conclusion

This study successfully cultivated bacteria from tissue obtained from the oesophagogastric tract and saliva. The risk of culture bias may select for favourable groups of bacteria and eliminate fastidious species. More important than the abundance, is understanding the functional capability of the individual bacteria to provide metabolic data pertinent to cancer pathogenesis. In order to do this, stable isotope labelling is needed to trace the active intrinsic pathways contributing to the metabolite pool.

5.2 PRE-CULTURE 16S rRNA GENE SEQUENCING ANALYSIS TO ASSESS CULTURE BIAS

5.2.1 Introduction

The upper gastrointestinal tract is considered 'sterile' owing to its highly acidic endoluminal environment. This poses an experimental challenge to extract adequate mucosal bacteria to enable accurate identification of microbial species. Low biomass samples are prevalent in other body system such as the lung, liver and biliary tract. To overcome this problem, a DNA extraction method was developed to optimise the DNA yield. DNA extraction kits are sources of potential contamination in low biomass samples. Other sources include reagents and materials used. The CTAB protocol was shown to increase nucleic acid yields by 30%.

5.2.2 Aims

The primary aim of this study is to characterise the oesophagogastric cancer and oral microbiome to assess the bias of culture methods using a novel technique.

5.2.3 Methods

Patient selection and Sample preparation, collection and storage are detailed in Section 2, Chapter 4.

Sample preparation, collection and storage of saliva All patients recruited remained fasted for greater than 6 hours as per hospital pre-procedure protocol. Participants provided a minimum of 1mL of whole saliva by passive drool into a sterile container. All samples were provided in the morning and patients adhered to their normal oral hygiene routines, at least 2 hours from sampling time. Saliva samples were aliquoted into cryovials and stored at -80°C before final processing.

DNA extraction Pre-culture mucosal bacteria derived from patient tissue biopsy and saliva samples were stored in PBS at -80°C. An established protocol (Department of Genomic Medicine, Brompton Campus, Imperial College London) was used to extract DNA from the low biomass samples. Hexadecyltrimethylammonium bromide, also known as CTAB (10% CTAB (w/v): phosphate buffer pH 8.0) was added to a maximum volume of 200µL of the sample in a lysing matrix E (LME) tube and incubated at room temperature for 15 minutes. 500µL phenol chloroform isoamyl alcohol (P:C:I) (25:24:1) was added to the LME tube and cell lysis performed in a bead beater (Next Advance Bullet Blender® Storm 24) with the following settings: speed 5.5m/second, 60 seconds, one cycle. After centrifugation at 16,000 g for 5 minutes, all liquid was transferred into the phase lock gel tube (2mL heavy duty). The LME was kept on wet ice for the second DNA extraction. The phase lock gel tube was centrifuged at 16,000 g for 5 minutes at 4°C. One volume of chloroform isoamyl alcohol (C:I) (24:1) was added and centrifuged again. These steps were repeated using the LME tube. Each sample had two resulting extraction tubes. The aqueous phase of both phase lock gel tubes were transferred to different

microcentrifuge tubes prepared with 1 μ L of linear polyacrylamide (LPA) and 2 volumes of 30% (w/v) polyethylene glycol 6000/1.6M sodium chloride (PEG/NaCl) solution. Samples were incubated overnight at 4°C to precipitate. The following morning, all samples were centrifuged at 16,000g for 20 minutes at 4°C. The PEG/NaCl solution was aspirated from the large translucent pellet. Pellets were washed with cold 70% ethanol, centrifuged at 16,000g for 5 minutes and supernatant discarded. After two further 200 μ L 70% ethanol washes, the pellets were air-dried for 5 minutes. The pellet was resuspended in 30 μ L of low EDTA TE buffer and combined with its counterpart to form a single tube of 60 μ L per sample. All samples were stored at -80°C until further processing. DNA was quantified using Qubit dsDNA high-sensitivity assay (Invitrogen™) with a range between 0.45 to 162.6 ng/ μ L.

PCR, Illumina MiSeq sequencing and Microbial data analysis DNA was amplified according to the Illumina protocol and sequencing on the MiSeq platform as detailed in Chapter 5.1. Microbial diversity and taxonomic data analysis was performed as per the DADA2 pipeline detailed in Chapter 5.1.

5.2.4 Results

Oesophagogastric cancer tissue microbiome Significant alpha diversity metrics were demonstrated between gastric cancer bacteria compared with normal stomach from healthy controls. Shannon index showed a significantly lower alpha diversity in gastric cancer samples from class to species taxonomic levels ($p < 0.05$). No alpha diversity differences were observed in oesophageal cancer bacteria samples. Beta diversity also showed no differences in both cohorts (Table 12). A novel panel of bacteria were identified using the chloroform DNA extraction method. *Patescibacteria* phylum was significantly reduced in both gastric and oesophageal cancer groups compared with controls ($p = 0.04$). This was observed for all taxonomic levels including *Saccharimonadia* (class; $p = 0.04$), *Saccharimondales* (order; $p = 0.04$), *Saccharimonadaeaceae* (family; $p = 0.04$) and *Saccharibacteria* TM7 (genus; $p = 0.04$). No other significances were found.

Saliva microbiome Saliva was obtained from oesophagogastric cancer patients ($n = 8$) and healthy controls ($n = 8$). Simpson diversity was significantly different in cancer saliva bacteria compared to healthy controls ($p = 0.04$). The remaining alpha and beta diversity metrics demonstrated no changes. *Actinobacteria* phylum was significantly higher in oesophagogastric cancer saliva ($p = 0.01$) and at all taxonomic levels down to *Actinomyces* genus ($p = 0.04$). *Solobacterium* ($p = 0.03$) and its species *Solobacterium moorei* ($p = 0.04$) was found to be significantly reduced in oesophagogastric cancer saliva. Another interesting finding was the high abundance of *Rothia* ($p = 0.03$) in cancer saliva (Figure 19).

Table 12. Alpha and beta diversity metrics (p value) for cancer bacteria samples compared to healthy controls.

	Alpha diversity					Beta diversity		
	Chao 1	ACE	Shannon	Simpson	Fisher	Bray-Curtis	Jensen-Shannon	Jaccard Index
SALIVA								
Oesophagogastric cancer vs Control								
Phylum	0.19	1.00	0.57	0.51	0.06	0.19	0.08	0.18
Class	0.23	0.47	0.57	0.44	0.08	0.23	0.10	0.25
Order	0.22	0.43	0.57	0.57	0.13	0.12	0.08	0.16
Family	0.22	0.32	0.19	0.19	0.06	0.15	0.22	0.19
Genus	0.37	0.37	0.28	0.13	0.28	0.29	0.32	0.35
Species	0.56	0.56	0.10	0.04	0.38	0.29	0.35	0.29
TISSUE								
Gastric cancer vs Control NS								
Phylum	0.44	0.35	0.18	0.33	0.13	0.15	0.25	0.18
Class	0.34	0.35	0.03	0.08	0.03	0.08	0.30	0.10
Order	0.23	0.06	0.02	0.05	0.02	0.22	0.32	0.19
Family	0.36	0.36	0.02	0.02	0.02	0.28	0.20	0.22
Genus	0.33	0.33	0.02	0.02	0.01	0.26	0.22	0.20
Species	0.36	0.36	0.02	0.05	0.13	0.34	0.30	0.29
Oesophageal cancer vs Control NO								
Phylum	0.36	0.68	0.76	0.76	0.20	0.06	0.44	0.07
Class	0.21	0.58	0.64	0.53	0.15	0.30	0.44	0.26
Order	0.33	0.41	0.64	0.88	0.34	0.23	0.28	0.17
Family	0.19	0.23	0.88	0.76	0.34	0.29	0.51	0.24
Genus	0.29	0.29	1.00	0.76	0.20	0.20	0.44	0.15
Species	0.46	0.46	0.88	0.88	0.20	0.30	0.50	0.22

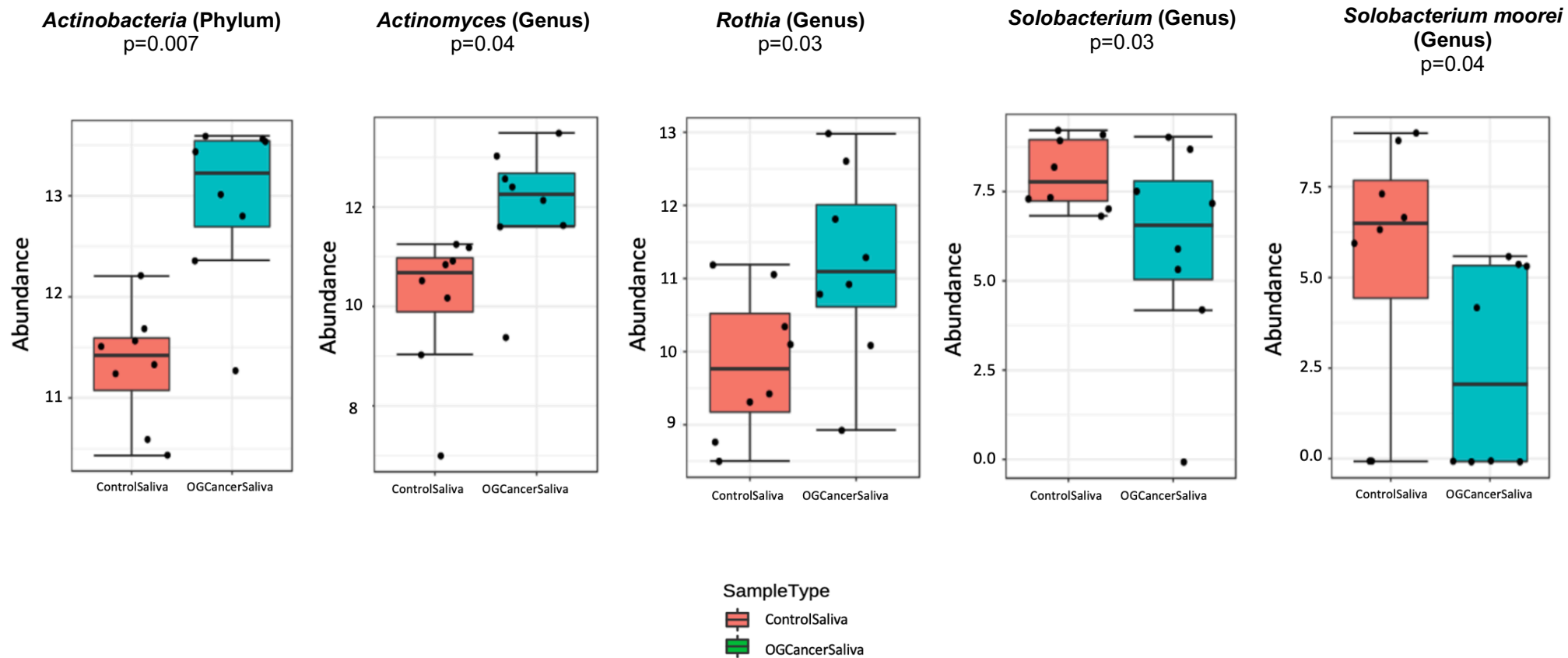


Figure 19. Box plots demonstrating select bacteria enriched and depleted in oesophageal cancer saliva compared to healthy control saliva. P value < 0.05 was considered statistically significant.

5.2.5 Discussion

The primary novel findings of this study were: (i) identification of cancer-specific salivary bacteria; namely *Actinomyces* and *Rothia* (ii) identification of novel human microbes and (iii) methodological bias.

This study has found high abundances of *Actinomyces* in the saliva of oesophagogastric cancer patients. This has been reported within the saliva of colorectal cancer cohorts in literature.[206] It is an anaerobic, fastidious, and opportunistic bacterium residing in the oral cavity. Conversely, *Solobacterium* has been identified as part of the normal salivary bacteriome, however has been found to be protective in cancer states. In keeping with these findings, reduced abundance of *Solobacterium* has also been linked with colorectal cancer. Aerobic *Rothia* is prevalent in human saliva and linked to gut disease processes such as gastric atrophy and intestinal metaplasia.[206] The findings corroborate with literature with significantly higher abundances of *Rothia* in cancer saliva.

The *Patescibacteria* group of bacteria have not convincingly been linked with human disease. It has been associated with saliva, skin and in an environmental context. In this study it was found to be significantly reduced in gastric and oesophageal cancer. The significance of this finding and role of this bacteria needs to be further defined.

The microbial profile identified with this technique differed from the previous method described in Section 5.1. A different DNA extraction method was chosen on this occasion to increase the yield of DNA. The chloroform method has been shown to increase DNA yield by 30%. However, this method is also not commonly used in the current era due to the toxicities associated with the reagents. Commercial kits are preferred, with the caveat they introduce contaminants. For high biomass samples such as saliva or faeces this doesn't present a problem, however in low biomass samples such as the upper gastrointestinal tract, contaminants can be amplified. These findings were not in keeping with literature, however, may warrant further exploration and be of particular interest in the scientific community. Another influencing factor was the method of bacteria collection. For this study, mucosal bacteria were separated from the tissue by centrifugation compared to the previous method of mechanically disrupting the tissue. Although this may introduce higher quantities of genomic DNA, gentle vortexing to isolate mucosal bacteria may not have been sufficient. These are methodological considerations for future work.

5.2.6 Conclusion

Cancer-specific bacteria has been identified in saliva, namely *Rothia* and *Actinomyces*, supported by literature associations with other gastrointestinal cancers. A reasonable comparison could not be made between pre-culture bacteria samples and cultured bacteria to determine potential bias of culture conditions. However, this study brings to light crucial methodological considerations for future studies.

SECTION 3

ASSESSMENT OF VOLATILE ORGANIC COMPOUNDS ASSOCIATED WITH THE ORAL MICROBIOME IN OESOPHAGOGASTRIC CANCER

6. CHAPTER 6: ASSESSING THE IMPACT OF ORAL CLEANSING STRATEGIES ON EXHALED VOLATILE ORGANIC COMPOUNDS

6.1 Introduction

The detection of volatile organic compounds (VOCs) within exhaled breath potentially offers a non-invasive approach for the detection and surveillance of human disease. An understanding of how VOCs are produced and released from the human body is critical to realising the diagnostic potential of breath. Specifically, such knowledge will support clearer recommendations regarding the requirement for standardisation of methods of breath collection in clinical practice.

Earlier work has confirmed the bacterial distribution in oesophagogastric cancer by isolating bacteria from cancer tissue and saliva. Literature findings have demonstrated the potential of the oral microbiome as a screening tool. However, this is the first time a potential relationship between the oral and cancer tissue microbiome has been suggested. In view of this, it was crucial to explore the role of oral hygiene and VOCs generated from within the oral cavity to help design future clinical studies.

In the majority of cases breath samples are acquired from an oral exhalation. The contribution of the oral cavity itself to the release of VOCs and the effect of oral hygiene is largely unknown in the context of breath research. The oral cavity serves many functions but principally acts as a route of entry to the aerodigestive tract. The oral cavity contains a large and diverse microbiome that itself can be an important source of VOCs.[207-209] Over 100 species inhabit the oral cavity, primarily residing on the dorsum of the tongue, in the saliva and between the teeth.[207, 208] Bacterial degradation of sulphur-containing amino acids cysteine and methionine produce volatile sulphur compounds.[208, 210] Less commonly, oral microbial metabolism produces volatile short chain fatty acids, polyamines and indoles.[209, 211] An increase in the production of VOCs within the oral cavity may arise from dietary contamination and/or poor oral hygiene and an associated increase in bacterial activity.[212, 213] Other factors affecting the oral environment include reduced salivary flow rates during sleep and inadequate water consumption.[207, 208]

6.2 Aim

The primary aim of this study is to assess the impact of oral contamination and subsequent cleansing measures on levels of VOCs detected within exhaled breath.

6.3 Methods

Participant selection The number of healthy participants recruited for this study was determined pragmatically. All participants had regular dental appointments, had no prior diagnosis of oral disease,

and were deemed to be at low risk of periodontitis. Invited participants who had disclosed current, or recent (<12 months) oral disease were excluded. The oral cavity of all participants was inspected by a medical professional to exclude any overt dental disease such as caries and gingival disease. Informed written consent was obtained from all participants prior to breath sampling with full ethical approval: London-Bloomsbury Research Ethics Committee 18/LO/0078.

Breath sample collection and analysis Breath samples were collected from ten healthy volunteers following an overnight fast (>6 hours), after consuming a standard nutritional challenge (doughnut). Subjects were required to provide further breath samples immediately after the following interventions: (i) no oral cleansing; (ii) oral rinse with water; (iii) oral rinse with salt water; (iv) toothbrushing with toothpaste (Colgate Original; Colgate-Palmolive Ltd, Guildford, UK), and (v) oral rinse with alcohol free mouth wash (Colgate Plax, Colgate-Palmolive Ltd, Guildford, UK). Sequential breath sampling was performed over a one-hour period for each participant with an interval of 10 minutes between each intervention. We established a protocol with the simplest forms of cleansing at the start, followed by interventions which were likely to introduce contamination. The objective was to study the effect of oral contamination and therefore the first breath sample taken was from a contaminated oral cavity as a baseline to observe changes with oral cleansing. Contents of the nutritional challenge, toothpaste and mouthwash are provided in Table 13. At each study time point subjects were requested to provide three repeat breath samples from which mean VOCs concentrations were determined.

SIFT-mass spectrometry and data analysis Direct on-line analysis of target VOCs was achieved using SIFT-MS VoiceUltra 200, (Syft Technologies, Anatune, UK). Subjects were instructed to exhale directly into the inlet of the SIFT-MS instrument via a disposable mouthpiece following resting tidal inhalation. Patients spontaneously inhaled during their normal breathing pattern and exhaled into the SIFT-MS for analysis of exhaled breath. SIFT-MS allows real-time quantification of VOCs within the breath based on chemical ionisation that occurs between selected reagent ions and trace. Reagent ions (H_3O^+ , NO^+ and O_2^+) are produced by a microwave discharge ion source and individually selected by a quadrupole mass filter. Having been selected, reagent ions are injected to inert helium carrier gas that subsequently passes along a flow tube. A sample (i.e. breath) is added into the carrier gas/reagent ion swarm via a calibrated heated capillary. Trace gases within the sample and reagent ions react to create characteristic product ions. Downstream, product ions are separated detected in accordance with their mass-to-charge ratio (m/z). Where the kinetics of the ion-molecular reactions are accurately known for the flow tube parameters, the SIFT-MS instrument can accurately detect and quantify multiple trace gases simultaneously. The SIFT-MS instrument was subjected to automated daily validation cycles for data quality assurance, with an ambient operating temperature between 10 - 30°C.

Details of exhaled VOCs that were analysed in this study, including their chemical formula, SIFT-MS reagent and product ions are presented in Table 14. VOCs were chosen from five chemical classes, which have been shown in previous studies to be linked to human disease, including cancer.[214] Abundant endogenous compounds acetone and isoprene were also analysed together with the

common flavourant menthone. VOC concentrations were identified and determined by the in-built kinetics library based on ion reactions which have been extensively studied previously.[215, 216]

Table 13. Details of interventions.

Intervention		Content
Nutritional challenge	Doughnut	67g serving, sugars 23.2g, proteins 3.3g, fats 9.9g (saturated 4.6g), salt 0.26g.
Water		Water
Salt water		Water, salt (1 teaspoon, ~4g)
Toothpaste	Colgate Original	Glycerine, aqua, hydrated silica, sodium lauryl sulphate, arginine, aroma, cellulose gum, zinc oxide, poloxamer 407, zinc citrate, tetrasodium pyrophosphate, xanthan gum, benzyl alcohol, cocamidopropyl betaine, sodium fluoride, sodium saccharin, phosphoric acid, sucralose. Contains: sodium fluoride 1450 ppm/F
Mouthwash	Colgate Plax mouthwash (alcohol-free)	Aqua, glycerine, propylene glycol, sorbitol, poloxamer 407, aroma, cetylpyridinium chloride, potassium sorbate, sodium fluoride, sodium saccharin, menthol, CI 42051. Contains: sodium fluoride (225ppm/F). Does not contain ethyl alcohol

Table 14. Details of volatile organic compound analysed by selected ion flow tube mass spectrometry.

Compound	Formula	Reagent Ion	Product Ion m/z
Acetone	C ₃ H ₆ O	H ₃ O ⁺	59
Ammonia	NH ₃	O ₂ ⁺	17
Isoprene	C ₅ H ₈	NO ⁺	68
Menthone	C ₁₀ H ₁₈ O	NO ⁺	154
Short Chain Fatty Acids			
Acetic acid	CH ₃ COOH	NO ⁺	90
Propanoic acid	CH ₃ CH ₂ COOH	NO ⁺	104
Butanoic acid	C ₄ H ₈ O	NO ⁺	118
Pentanoic acid	C ₅ H ₁₀ O ₂	NO ⁺	85
Hexanoic acid	C ₆ H ₁₂ O ₂	NO ⁺	146
Alcohols			
Methanol	CH ₃ OH	H ₃ O ⁺	33
Ethanol	C ₂ H ₅ OH	H ₃ O ⁺	47
1-pentanol	C ₅ H ₁₂ O	H ₃ O ⁺	71
Aldehydes			
Acetaldehyde	C ₂ H ₄ O	H ₃ O ⁺	45
Propanal	C ₃ H ₆ O	NO ⁺	57
Butanal	C ₄ H ₈ O	NO ⁺	71
Pentanal	C ₅ H ₁₀ O	NO ⁺	85
Hexanal	C ₆ H ₁₂ O	NO ⁺	99
Heptanal	C ₇ H ₁₄ O	NO ⁺	113
Octanal	C ₈ H ₁₆ O	NO ⁺	127
Nonanal	C ₉ H ₁₈ O	NO ⁺	141
Phenols			
Phenol	C ₆ H ₅ OH	NO ⁺	94
P-cresol	C ₇ H ₈ O	NO ⁺	108
1-hydroxy-4-ethylbenzene	C ₈ H ₁₀ O	NO ⁺	122
Hydrocarbons			
Decane	C ₁₀ H ₂₂	NO ⁺	141
Dodecane	C ₁₂ H ₂₆	H ₃ O ⁺	189
Sulphur compounds			
Hydrogen sulphide	H ₂ S	H ₃ O ⁺	35
Methyl mercaptan	CH ₄ S	H ₃ O ⁺	49
Dimethyl sulphide	C ₂ H ₆ S	O ₂ ⁺	62
Dimethyl disulphide	C ₂ H ₆ S ₂	NO ⁺	94
Dimethyl trisulphide	C ₂ H ₆ S ₃	H ₃ O ⁺	127

6.4 Results

Ten healthy participants (7 female; 28 ± 6 years) provided breath samples after consuming a standard nutritional challenge and subsequent oral cleansing measures. Median concentrations of measured VOCs are summarised in Table 15.

Acetone, isoprene and menthone

Levels of acetone and isoprene did not vary significantly as a consequence of oral cleansing, confirming their systemic origin (Figure 20).[217] Proposed reaction suggested menthone levels largely increased after toothbrushing and mouthwash use reflecting the flavoured ingredient within the products. Note that the menthone product ion (m/z 154) overlaps with decanal product ion (m/z 155).

Short chain fatty acids

The levels of exhaled butanoic and pentanoic acid declined after rinsing with water. Whilst pentanoic acid levels did not appear to decline further with subsequent oral cleansing with saltwater, tooth brushing and mouthwash, these interventions led to an incremental decline in butanoic acid levels. Methods designed to achieve oral cleansing did not appear to significantly influence the levels of acetic-, hexanoic- and propanoic acid. In addition, the reducing variability across subjects is consistent with the oral cleansing measures, except an increase observed with saltwater (Figure 20).

Alcohols

Ethanol levels fell significantly after rinsing with water and thereafter remained relatively unchanged despite further interventions. After rinsing with water, 1-pentanol levels were significantly decreased with only modest benefit from further cleansing measures. Methanol levels did not vary significantly as a result of oral cleansing. A consistent decline in variability was seen with maximal effect after an oral water rinse.

Aldehydes

It is suggested that acetaldehyde levels declined significantly after rinsing with water, whilst subsequent cleansing with saltwater, toothbrushing and mouthwash did not appear to affect the measured concentration of this VOC. Propanal, hexanal, heptanal and nonanal levels were elevated after rinsing with mouthwash. Heptanal was also elevated after toothbrushing. Variability among subjects remained unchanged until toothbrushing and mouth washing intervention. Acetaldehyde was the exception and showed a similar pattern to the alcohol group.

Phenols and Hydrocarbons

Decane, dodecane and P-cresol were increased in breath after toothbrushing and rinsing with mouthwash. Phenol levels were lowest after tooth brushing, whilst 1-hydroxy-4-ethylbenzene levels were not significantly altered by any of the four interventions. The overall trend in variability followed that of aldehydes with an increase in after toothbrushing and mouth washing. Ammonia, a known

product of mouth microbiota, progressively declined with each successive intervention, maximally with the oral water rinse.

Sulphur compounds

Reactions suggested dimethyl disulphide levels fell by the greatest amount after toothbrushing with a moderate rise after rinsing with mouthwash. Dimethyl sulphide increases after rinsing with mouthwash. Other sulphur compounds were unaffected by oral cleansing measures. No consistency of subject variability was observed.

Table 15. Variation in selected volatile organic compound concentrations (ppbv) in response to oral cleansing measures.

	After nutrient challenge (ppbv)	Water rinse (ppbv)	Saltwater rinse (ppbv)	Toothbrush (ppbv)	Mouthwash rinse (ppbv)	P*
Acetone	721.5 [511.8-997.1]	705.5 [464.3-979.4]	742.5 [513.8-914.3]	733.0 [503.7-949.9]	671.7 [506.5-896.8]	0.988
Ammonia	332.2 [284.4-587.2]	247.4 [169.0-383.7]	230.3 [170.3-477.2]	209.8 [183.8-390.4]	199.5 [172.3-403.5]	0.004
Isoprene	99.4 [74.9-104.0]	106.0 [73.4-140.0]	107.0 [79.7-124.6]	95.3 [80.4-105.1]	97.3 [84.5-106.1]	0.904
Menthone	0.4 [0.3-0.4]	0.3 [0.2-0.5]	0.3 [0.3-0.5]	185.3 [119.1-248.9]	225.5 [145.8-332.1]	<0.001
Short Chain Fatty Acids						
Acetic acid	26.7 [20.5-32.8]	18.7 [17.6-23.9]	14.9 [12.9-25.0]	18.6 [15.7-25.7]	21.4 [18.7-27.1]	0.237
Butanoic acid	13.1 [10.6-14.9]	9.0 [4.8-14.4]	6.4 [2.9-10.1]	4.0 [3.5-4.9]	3.5 [2.9-4.1]	<0.001
Hexanoic acid	0.7 [0.7-0.9]	0.8 [0.6-1.0]	0.8 [0.6-0.9]	0.8 [0.7-1.0]	0.7 [0.7-0.8]	0.184
Pentanoic acid	2.4 [1.9-2.5]	1.2 [1.1-1.4]	1.2 [1.1-1.6]	1.7 [1.2-1.9]	1.7 [1.7-2.1]	<0.001
Propanoic acid	11.9 [9.2-13.1]	11.1 [7.7-12.9]	9.1 [6.3-11.1]	9.3 [7.3-11.2]	10.2 [8.2-13.3]	0.291
Alcohols						
Methanol	299.5 [146.4-337.1]	281.8 [147.2-369.5]	303.7 [227.7-391.1]	328.2 [277.8-397.6]	287.5 [252.0-374.8]	0.126
Ethanol	315.7 [205.2-455.0]	61.2 [44.6-78.3]	58.0 [40.9-71.4]	63.1 [44.9-100.7]	63.8 [50.0-107.6]	0.003
1-pentanol	2.3 [2.1-2.5]	1.2 [1.1-1.7]	1.1 [0.9-1.2]	1.3 [1.1-1.6]	0.9 [0.8-1.0]	<0.001
Aldehydes						
Acetaldehyde	21.0 [19.4-23.2]	15.5 [12.8-19.6]	16.3 [12.5-18.1]	17.6 [14.1-22.1]	17.8 [14.5-18.7]	0.009
Heptanal	0.6 [0.5-0.8]	0.5 [0.5-0.6]	0.6 [0.5-0.6]	0.8 [0.8-0.9]	1.8 [1.3-2.3]	<0.001
Hexanal	2.8 [2.5-3.0]	2.2 [1.9-2.3]	2.0 [1.9-2.1]	2.8 [2.5-3.3]	3.1 [2.6-3.3]	0.001
Nonanal	0.7 [0.6-0.7]	0.7 [0.6-0.8]	0.7 [0.6-0.7]	1.1 [0.8-1.3]	1.3 [1.1-1.6]	<0.001
Octanal	0.6 [0.5-0.6]	0.6 [0.5-0.6]	0.6 [0.5-0.6]	0.8 [0.8-1.0]	0.5 [0.5-0.7]	0.069
Pentanal	0.5 [0.4-0.6]	0.6 [0.5-0.7]	0.5 [0.5-0.6]	0.7 [0.5-0.7]	0.7 [0.5-0.8]	0.308
Butanal	1.3 [1.1-1.7]	1.1 [1.0-1.4]	1.3 [1.1-1.6]	1.5 [1.1-1.6]	1.3 [1.1-1.4]	0.512
Propanal	19.1 [15.9-56.9]	18.7 [15.0-54.2]	18.3 [16.6-52.7]	18.8 [16.9-57.7]	24.3 [20.2-61.4]	0.001
Phenols						
Phenol	3.1 [1.9-4.9]	3.1 [1.7-3.6]	3.2 [2.4-3.9]	2.1 [1.8-3.2]	4.0 [3.0-4.5]	0.001
P-cresol	1.3 [1.2-1.9]	1.3 [1.3-1.5]	1.3 [1.2-1.7]	2.0 [1.6-3.1]	1.9 [1.6-2.6]	<0.001
1-hydroxy-4-ethylbenzene	0.9 [0.6-1.0]	0.8 [0.6-0.9]	0.7 [0.6-0.8]	0.7 [0.6-2.1]	0.7 [0.6-1.4]	0.690
Hydrocarbons						
Decane	1.3 [1.2-1.5]	1.4 [1.1-1.4]	1.3 [1.1-1.5]	2.1 [1.6-2.3]	2.4 [1.5-2.9]	0.005
Dodecane	0.7 [0.5-0.9]	0.7 [0.5-0.7]	0.6 [0.5-0.9]	1.2 [1.0-1.3]	1.3 [0.9-1.5]	0.001
Sulphur compounds						
Dimethyl disulphide	2.9 [1.7-3.4]	2.4 [1.7-3.2]	2.3 [2.3-3.8]	2.2 [1.7-2.5]	3.1 [2.6-4.1]	0.025
Dimethyl sulphide	12.9 [11.9-17.0]	11.8 [11.0-14.2]	11.9 [10.0-14.9]	12.7 [10.2-14.6]	19.4 [15.8-22.5]	<0.001
Dimethyl trisulphide	2.0 [1.7-2.3]	1.9 [1.7-2.2]	2.1 [1.9-2.1]	2.1 [1.6-2.5]	1.8 [1.7-2.4]	0.796
Hydrogen sulphide	17.6 [10.6-33.5]	12.6 [6.1-22.4]	11.8 [7.5-27.7]	16.5 [8.3-25.4]	16.1 [7.5-26.9]	0.054
Methyl mercaptan	1.1 [0.9-1.3]	0.8 [0.5-0.9]	0.9 [0.7-1.1]	0.9 [0.6-1.1]	0.9 [0.9-1.2]	0.479

Values are presented as median [IQR], * Friedmans test

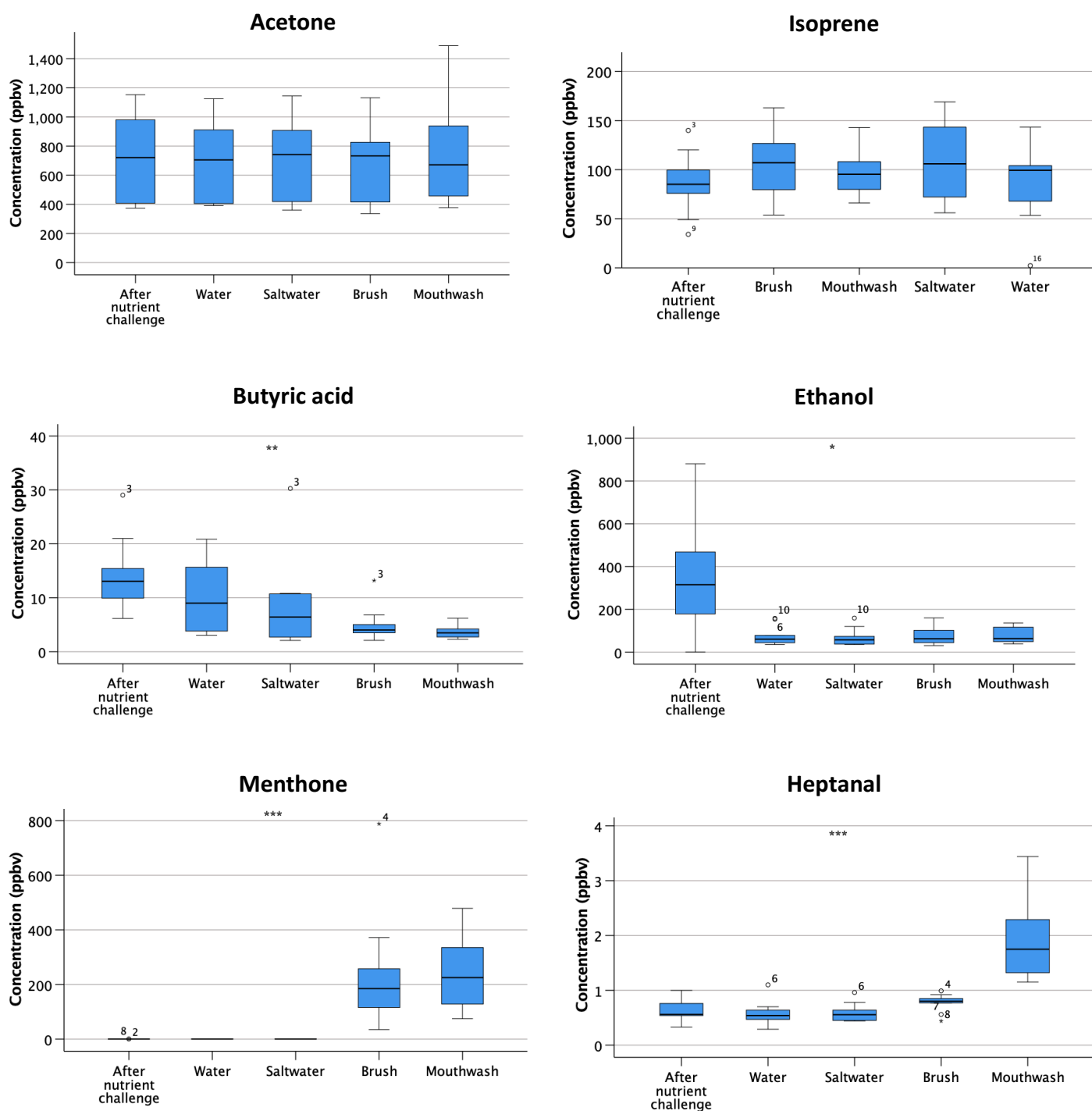


Figure 20. Variation in volatile organic compounds in response to oral cleansing measures. There is minimal variation in acetone and isoprene suggesting their systemic origin. Levels of butyric acid and ethanol reduce in response to oral cleansing. Heptanal and menthone are a source of contamination from oral cleansing ingredients.

6.5 Discussion

This study systematically assesses the effect of oral cleansing measures on the concentrations of clinically relevant VOCs within exhaled breath. The principal finding of this study was that whilst oral cleansing is not necessary for all VOCs, a simple oral rinse with water prior to breath testing may partially mitigate the effects of oral contamination and minimise variability of the baseline among subjects. We used an incremental method of oral cleansing, and in theory there may be an additive effect, however, the results demonstrate no significant changes after the oral water rinse, except after the use of toothpaste and mouthwash. This was expected due to the volatile ingredients present in the products. Findings are intended to inform future practice in clinical trials involving breath analysis.

An important finding of this study suggests that acetone and isoprene levels are not significantly altered by oral cleansing measures. Both acetone and isoprene are found in relative abundance (even distribution across individuals) within breath and are recognised to be primarily of systemic origin. Results verify that the oral cavity is not likely to contribute significantly to the levels of these VOCs within breath, hence providing an important negative control for chosen methods of oral cleansing. With the exception of butyric acid there is limited evidence supporting the use of further cleansing strategies. Elevated levels of selected aldehydes, phenols and sulphur compounds suggest that the use of toothpaste and mouthwash may contaminate exhaled breath. Previous work by our group suggests aldehydes, in particular decanal, is a potential biomarker for early detection of oesophagogastric cancer.[214] The overlap between decanal and menthone is an important finding as use of these interventions containing flavoured ingredients may compromise results. Rinsing with water prior to breath sampling is a simple intervention which may reduce subject specific variability associated with exhaled VOC analysis. The advantages of patients not having to fast prior to breath sampling include greater potential for opportunistic testing in future clinical practice, and improved patient acceptability.

An alternative method to bypass contamination from the oral microbiota is nose exhalation, however, this presents its own challenges. Direct sampling via the SIFT requires a sample line to be inserted in the nostril, however this is associated with (i) loss of breath either side of the sample line (ii) discomfort (iii) variability in results due to the quality of the nose breath sample. Optimal nose breath collection would require a fitted nose mask; however, this is currently not available.

It is recognised that the nutritional challenge contained sugars, fats, and proteins, however, this study did not seek to determine the effects of specific substrates on breath VOC levels. It is important to note that diets that are rich in specific substrates may result in the disproportion production of VOCs by the nature of their metabolic by-products.

6.6 Conclusion

This study concludes that rinsing with water reduces oral contamination VOC levels, and equally that further attempts of oral decontamination using flavoured products may adversely affect results. This is a simple and inexpensive intervention which may serve as an important method of standardisation within breath research potentially reducing the requirement for prolonged fasting. It also eliminates a dry mouth from reduced salivary flow rates which influences the VOC production. All future breath studies will employ oral water rinsing prior to breath sampling.

7. CHAPTER 7: METHODOLOGICAL OPTIMISATION OF SALIVARY HEADSPACE VOC ANALYSIS

7.1 METHODOLOGICAL OPTIMISATION OF SALIVARY HEADSPACE VOC ANALYSIS

7.1.1 Introduction

The previous study has highlighted the importance of the role of the oral cavity in VOC production. This may be multifactorial, incorporating saliva, tongue coating and dental plaques. Based on the potential relationship demonstrated between the saliva and cancer tissue microbiome, this study investigated VOCs obtained directly from the headspace of saliva. Saliva offers a potentially non-invasive tool for early disease diagnostics. Saliva is a complex liquid composed of electrolytes, enzymes and mucin suspended in water.[218] The molecular composition of saliva is altered with autonomic stimulation.[218, 219] It is important to acknowledge there may be compositional differences between unstimulated and stimulated saliva. Salivary flow rates can influence this, as unstimulated saliva flow rate is 0.3-0.4 mL/min and stimulated flow rates can reach up to 4 mL/min.[220] There is a relative paucity in studies reporting salivary VOCs analysis, where much of the research exists in salivary proteomics.[221] Factors affecting saliva VOCs may include: (i) local factors such as oral hygiene, oral disease, smoking status and (ii) systemic factors such as significant systemic disease, diet and use of medications. Earlier studies in healthy controls suggested variability in day-to-day saliva sampling with considerable inter- and intra-individual variability.[222]

7.1.2 Aim

The aim of this study is to develop an optimised method for standardise salivary VOC analysis.

7.1.3 Methods

Participant selection Healthy participants were recruited from Imperial College London from February 2020 to March 2020. Inclusion criteria were healthy controls aged over 18 years and able to provide informed consent. Exclusion criteria included: current smoker, known oral disease, severe systemic disease, use of immunosuppressive medications. All participants provided written informed consent. Ethical approval was obtained by the Health Research Authority National Research Ethics Service (NRES) committee London Camden and Islington; REC number: 14/LO/1136.

All saliva sampling was performed between 0700-0900 after following a clear fluid diet for 6 hours. Normal oral hygiene routines were maintained with a minimum of one hour before saliva sampling. Salivary flow rate was calculated according to the following equation:

$$\text{salivary flow rate (mL/minute)} = \text{volume (mL)} / \text{time (minutes)}$$

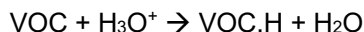
Saliva sampling parameters Four sampling parameters were assessed to optimise saliva headspace methodology: (i) saliva collection (ii) saliva volume (iii) saliva dilution and (iv) saliva acidification.

- (i) Saliva collection: Three techniques were employed (a) passive drool into a sterile plastic container over 5 minutes (b) unstimulated saliva using a Salivette® (c) stimulated saliva using a Salivette®. Salivette® (Sarstedt, Nümbrecht) is a commercially available saliva collection device composed of a cotton roll within a plastic collection tube suitable for centrifugation. Unstimulated saliva was obtained by holding the cotton roll in the mouth for two minutes. Acceptable sites included between the lateral teeth and buccal mucosa, or beneath the tongue. Stimulated saliva was collected by chewing on the cotton roll for one minute. Subsequently, the roll was placed into the collection tube and centrifuged for two minutes at 1000 g, according to manufacturer's instructions. The osmolality of saliva was determined for each method with 50µL of saliva using an osmometer (Advanced Instruments OsmoTech Promulti-Sample Micro Osmometer).
- (ii) Saliva volume: Volumes of 500µL and 1000µL of saliva were chosen based on literature findings.
- (iii) Saliva dilution: 500µL of whole saliva was compared with saliva diluted to 1mL with MiliQ water. Increasing the aqueous volume will reduce the gaseous volume and theoretically concentrate the headspace.
- (iv) Saliva acidification: A standardised protocol for measuring pH levels was used (FiveGo™ pH meter F2, Mettler Toledo). Calibration checks were performed with standard buffer solutions of pH 4.0 and pH 7.0. Each sample was acidified with pure hydrochloric acid with increasing increments of 0.1µL until pH 2.0 was achieved. An average of 3.7µL hydrochloric acid was used (range 3-5µL).

Control studies A baseline of VOCs emitted from the equipment was established to define the contribution of background VOCs to the final composition. Triplicate controls included: (i) a dry cotton roll from the Salivette® (ii) a cotton roll saturated with MiliQ water for 2 minutes (iii) a cotton roll saturated with MiliQ water followed by centrifugation (iv) headspace vials containing 500µL, 1000µL and 1500µL of MiliQ water.

Sample processing and analysis All samples were stored at 4°C and analysed within 2 hours of collection. All samples were aliquoted into 20mL glass headspace vials sealed with 18mm metal screw-tops and a polytetrafluoroethylene (PTFE) septum. Vials were incubated for 30 minutes at 37°C in a hot water bath to reflect the physiological body temperature. Septa were pierced with a sterile blunt needle (18G, 1 ½ (1.2mm x 40mm)) with real-time headspace VOC monitoring using the proton transfer reaction-time of flight-mass spectrometer (Ionicon Analytik GmbH Innsbruck, Austria). Twenty seconds of headspace sampling was followed by a 10 second interval.

Proton Transfer Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS) A direct untargeted analysis was performed. The hydronium ion (H_3O^+) was selected as the primary reagent ion. The PTR-ToF-MS functions on the principle of proton transfer from the hydronium ion due to the proton affinity of volatile compounds being greater than that of water.



Hydronium ions are formed from humidified air within the ionisation chamber. The sample is taken into the drift tube with an inlet flow set to 40 sccm. Electric fields propel the ions into the drift tube to interact with the injected sample. The drift tube parameters are optimised at: temperature 110°C, pressure 2.29 mbar, voltage 350V, resulting in an E/N of 84 Td (E=electric field, N=gas number density). Soft chemical ionisation reduces fragmentation rates. Following protonation of the trace analytes, electric fields direct ions from the reaction chamber towards the time-of-flight (ToF) detector. The ToF propels the analytes according to their mass, with heavier ions travelling at a slower pace. The detector identifies VOCs according to their mass-to-charge ratio (m/z) simultaneously and determines concentrations according to signal intensity. The PTR-ToF-MS cannot annotate compounds directly, however specific molecular weights can tentatively be assigned a compound. Quality control checks were employed daily for characterisation of impurities, accuracy, fragmentation rates, and resolution of the three reagent ions. Select compounds from a permeation unit (ES 4050P, Eso Scientific, Stroud, Gloucestershire UK) provides a steady flow of acetone, benzene, and phenol directly into the PTR-ToF-MS. Direct analysis offers the ease of no sample preparation which may result in VOC loss.

Data analysis PTR-ToF-MS data was extracted using PTR-MS Viewer version 3.2.8.0 (Ionicon Analytik) and analysed using an in-house generated R script. Descriptive statistics were used to present the data.

7.1.4 Results

Twelve participants were recruited with a median age of 25 years (IQR: 20-29) and 7 (58%) were female. Initial control experiments were conducted to assess the contribution of VOCs from the saliva collection equipment. The salivette® cotton roll demonstrated a high release of all SCFA (acetic acid 184 ppbv, butyric acid 23 ppbv) and acetone (3649 ppbv). The cotton roll soaked with MiliQ water also showed high levels of all SCFA, albeit lower than the dry cotton roll. Acetone was significantly reduced to 444 ppbv from 3649 ppbv. Following centrifugation, the analysis of the supernatant showed further reduction of VOC concentrations. Acetic acid reduced from 184 ppbv to 37 ppbv, and butyric acid from 23 ppbv to 3.1 ppbv. This was consistent across all SCFA assessed, suggesting the cotton roll within the salivette® was a significant source of contamination. There was a wide coefficient of variation between triplicate samples highlighting the lack of reproducibility with the use of the salivette® as a saliva collection tool (Table 16). Three different volumes of MiliQ water were tested to assess the impact

of volume on VOCs detected. Across all compounds of interest, these remained consistent with a low coefficient of variation (<10%).

Osmolality and flow rates of saliva were assessed to identify the differences between three saliva collection methods. Sequential saliva samples were provided from the least stimulated method to the stimulated method. Passive drool samples showed a higher osmolality (99 mOsm) compared to stimulated samples (64 mOsm). Salivary flow rates also increased with stimulation. Passive drool and unstimulated saliva had flow rates of 0.7 mL/min which increased to 1.9 mL/min with stimulation (Table 17).

Saliva collection Both passive drool and unstimulated saliva methods yielded comparable concentrations of acetone and SCFA of interest; acetic acid 14 vs 14 ppbv, butyric acid 2.5 vs 3.2 ppbv, respectively. Stimulated saliva generated up to 2-fold higher VOC concentrations.

Saliva volume Two volumes of 500µL and 1000µL were assessed using the three methods of collecting saliva. Passive drool produced comparable VOC concentrations between both volumes with negligible variability (butyric acid: 2.5 ppbv vs 2.3 ppbv). 500µL of unstimulated and stimulated saliva using the salivette® generated 2-fold higher concentrations of VOC, with a large variability. This is suggestive of contributory effects from the salivette® cotton roll.

Saliva dilution Passive drool methods again showed little variability between saliva with and without dilution with MiliQ, and minimal variability between participating subjects (butyric acid: 2.5 ppbv vs 2.3 ppbv).

Saliva acidification Acidifying saliva to pH 2.0 did not increase the concentrations of volatile SCFA (butyric acid 16 ppbv vs 17 ppbv). (Table 18, Figure 21).

Table 16. Concentrations of VOC of interest in control experiments assessing the VOC contribution of saliva collection equipment.

	Acetone (m/z 59)	Acetic acid (m/z 61)	Butyric acid (m/z 89)	Pentanoic acid (m/z 103)	Hexanoic acid (m/z 117)
Salivette®					
Dry cotton roll	3649 (75)	184 (97)	23 (84)	6.6 (83)	5.0 (66)
Wet cotton roll (MiliQ)	444 (150)	134 (84)	11 (94)	3.7 (67)	1.6 (72)
Centrifuged wet cotton roll (MiliQ)	138 (107)	37 (11)	3.1 (39)	2.0 (15)	0.9 (31)
Volume					
500uL	15 (1.3)	12 (1.5)	1.4 (2.6)	0.7 (2.6)	0.9 (5.1)
1000uL	15 (2.2)	11 (6.0)	1.3 (2.3)	0.7 (3.8)	1.0 (2.2)
1500uL	14 (3.0)	10 (2.7)	1.2 (6.5)	0.6 (3.9)	0.7 (9.9)

mean concentration, ppbv (Coefficient of Variation, %)

Table 17. Osmolality (mOsm) and flow rates (mL/min) of saliva retrieved with three different collection methods.

	Osmolality (mOsm)	Flow rate (mL/min)
Passive drool	99 (80-131)	0.69 (0.52-0.90)
Unstimulated saliva	81 (70-85)	0.70 (0.57-1.07)
Stimulated saliva	64 (48-80)	1.89 (1.34-2.21)

mean concentration, ppbv (interquartile range)

Table 18. Volatile short chain fatty acid concentrations (ppbv) to assess the optimal analysis of saliva headspace by using various methodological parameters.

	Acetone (m/z 59)	Acetic acid (m/z 61)	Butyric acid (m/z 89)	Pentanoic acid (m/z 103)	Hexanoic acid (m/z 117)
Saliva collection					
Passive drool	60 (46-86)	14 (13-14)	2.5 (2.4-4.0)	1.0 (0.8-1.1)	0.8 (0.7-0.9)
Unstimulated saliva	104 (61-1793)	14 (14-37)	3.2 (1.8-11.1)	0.7 (0.6-3.2)	0.7 (0.6-2.3)
Stimulated saliva	866 (278-1365)	24 (16-27)	6.1 (2.6-9.8)	2.0 (0.8-2.3)	1.1 (0.8-1.5)
Volume					
Passive drool					
500uL	57 (46-90)	14 (13-14)	2.5 (2.3-3.9)	0.9 (0.8-1.1)	0.8 (0.7-0.9)
1000uL	51 (37-58)	13 (13-14)	2.3 (1.6-3.2)	0.8 (0.7-1.0)	0.8 (0.5-0.8)
Unstimulated saliva					
500uL	1430 (390-1883)	28 (16-44)	8.2 (4.0-12)	2.0 (0.9-3.9)	1.8 (0.8-2.3)
1000uL	581 (213-1538)	18 (15-44)	4.2 (3.3-12)	1.2 (0.8-4.3)	1.0 (0.7-2.4)
Stimulated saliva					
500uL	866 (453-1365)	24 (19-27)	6.1 (3.2-9.7)	2.0 (1.3-2.6)	1.3 (0.9-1.6)
1000uL	173 (71-946)	14 (13-26)	4.5 (1.9-6.5)	0.8 (0.6-2.9)	0.7 (0.6-1.6)
Dilution					
Passive drool					
500uL saliva	60 (46-86)	14 (13-14)	2.5 (2.4-4.0)	1.0 (0.8-1.1)	0.8 (0.7-0.9)
500uL saliva + 1mL MiliQ	41 (29-59)	13 (13-14)	2.3 (1.8-4.0)	0.9 (0.7-0.9)	0.8 (0.7-0.9)
Unstimulated saliva*					
500uL saliva	104 (55-104)	14 (14-14)	1.8 (1.6-3.2)	0.7 (0.6-0.7)	0.6 (0.6-0.7)
500uL saliva + 1mL MiliQ	43 (40-55)	14 (14-15)	1.7 (1.7-1.9)	0.7 (0.7-0.7)	0.6 (0.6-0.7)
Stimulated saliva†					
500uL saliva	1589	28	8.9	2.2	1.8
500uL saliva + 1mL MiliQ	512	19	4.0	1.3	1.0
Acidification					
Non-acidified	761 (635-1500)	85 (43-108)	16 (13-21)	3.9 (2.6-6.0)	2.0 (1.2-2.7)
Acidified (pH 2.0)	692 (556-932)	91 (47-160)	17 (10-22)	3.1 (2.6-5.7)	2.0 (1.2-2.5)

mean concentration, ppbv (Interquartile range, ppbv)

* 5 participants

† 3 participants

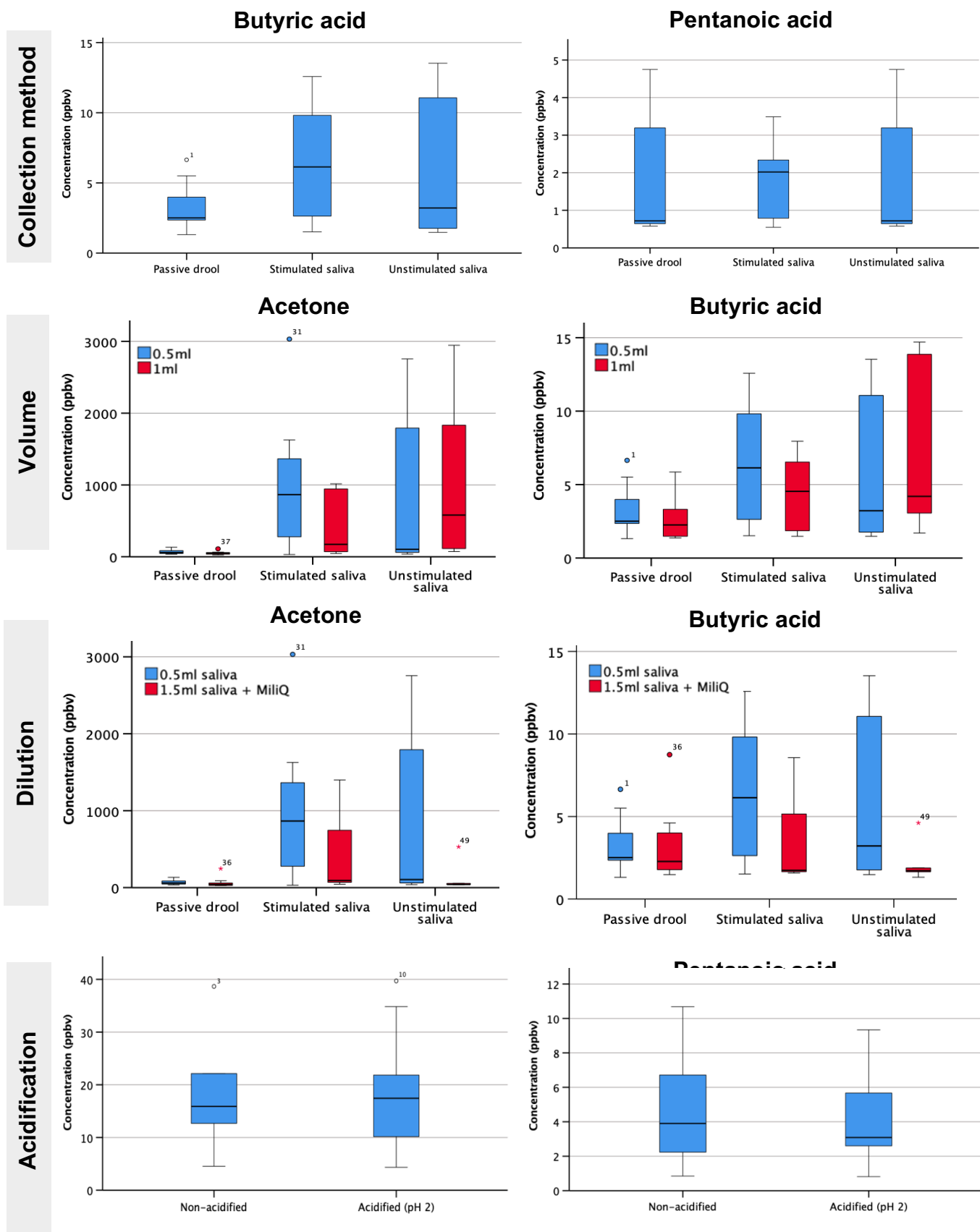


Figure 21. Four methodological parameters of saliva headspace optimised by VOC concentration (ppbv) assessment: saliva collection method, saliva volume, saliva dilution, saliva acidification.

7.1.5 Discussion

This study supports a simple, inexpensive and acceptable method for collection of saliva as a non-invasive investigative tool. The primary findings are: (i) saliva collected by the passive drool method presents the least contaminated collection method, with negligible variability between subjects (ii) a volume of 500 μ L is acceptable for headspace analysis (iii) no requirements for dilution or acidification of the sample. This presents the simplest form of saliva collection with no pre-processing steps which can introduce contamination.

The composition of saliva is influenced by interlinked factors such as flow rates and time of sampling. Unstimulated saliva flows at 0.3-0.4mL/min and is reduced by a further 0.1mL/min during sleep.[220] Therefore morning saliva samples following an overnight fast may be more concentrated with reduced flow. This was supported by our osmolality and flow rate assessments which showed a reduced osmolality and increased flow rate with stimulated saliva (64mOsm, 1.9mL/min) compared to passive drool (99mOsm, 0.7mL/min). Although passive drool and the unstimulated method yielded comparable results, the salivette® cotton roll introduced significant contamination of target compounds.

Although this study presents a standardised model for saliva collection, the lack of normalisation of the results have been considered. Saliva normalisation has largely been established in the field of proteomics research where osmolality or protein content, a measure of solute concentration within the given volume, can be used effectively. Consideration has been given to a variety of normalisation techniques: (i) measuring the concentrations of target VOCs against the total concentration of all VOCs detected (ii) measuring the concentration of a single target VOC against the total concentration of target VOCs; presented as ratios and/or percentage values. However, none of these strategies offered valuable information and were deemed appropriate. In general, normalisation of VOC concentrations is not routinely employed and therefore it was decided that if salivary flow rates and osmolality were within acceptable ranges, no further assessment was required.

7.1.6 Conclusion

Passive drool of 500 μ L of saliva is sufficient for headspace analysis of VOCs. Standardisation with morning sampling after a fasted state of 6 hours is recommended and supported by this study demonstrating minimal inter-subject variability. This technique will be translated into a clinical study to assess VOCs present in the saliva of patients with oesophagogastric cancer. Ultimately, saliva analysis potentially offers a non-invasive method for earlier cancer detection.

7.2 ASSESSMENT OF THE LONGITUDINAL VARIATION IN VOC LEVELS IN WHOLE SALIVA

7.2.1 Introduction

Volatile organic compounds (VOCs) are potential non-invasive biomarkers of disease. Variation in VOCs levels may reflect an individual's metabolic profile and can be detected within biological matrices such as breath, urine and saliva. Disease-specific salivary VOCs have been reported in benign (halitosis, oral abscess) and malignant (gastrointestinal cancer) states. Before wider adoption of salivary volatolomics as a non-invasive diagnostic strategy, it is first important to establish the variation in VOC profiles both within and between healthy subjects.

7.2.2 Aim

To assess the intra- and inter- individual variability of VOC levels within whole saliva.

7.2.3 Methods

Participant selection Healthy participants were recruited in July 2021. Inclusion criteria included: aged between 20-35 years, able to provide written informed consent. Exclusion criteria included: known oral disease, known severe systemic disease, use of immunosuppressive medications, active smokers, recent use of antibiotics within 4 weeks. All participants maintained their normal oral hygiene routines at least two hours prior to saliva sampling. Unstimulated saliva samples were provided on five mornings: three samples on week one, one sample in week two and one sample in week three. These were selected to assess the day-to-day and week-to-week variability. All participants provided written informed consent. Ethical approval was obtained via the Imperial College Healthcare Tissue Bank; REC Wales approval: 17/WA/0161 (R20059).

Sample processing A minimum of 3mL of passive drool saliva samples were collected into a sterile container and timed to calculate flow rates. Saliva was immediately aliquoted into 1.8mL cryovials and stored at -80°C until analysis. Prior to analysis all saliva samples were thawed at room temperature. 500µL was aliquoted in triplicates into 20mL glass headspace vials secured with 18mm metal screw-tops and PTFE septa. 20µL was analysed on the osmometer (Advanced Instruments OsmoTech Promulti-Sample Micro Osmometer) to measure osmolality of each sample. All samples were incubated at 37°C for 30 minutes in a water bath to equilibrate the VOCs between the aqueous and gaseous phases. Empty headspace vials were used as controls.

Mass spectrometry analysis Direct sampling was performed using the SIFT-MS with a targeted method for volatile short chain fatty acids (C2-C6) and aldehydes, previously established by the group. The septa were pierced with a sterile blunt needle (18G, 1 ½ (1.2mm x 40mm)) with an analytical time frame of 20 seconds and an inlet flow rate of 25mL/min.

Statistical analysis Friedman's test was used to identify statistically differences between the five sets of samples.

7.2.4 Results

Ten healthy participants (7 female) were recruited with a median age of 30 (IQR: 25-32). The mean salivary flow rate was 0.7ml/minute (IQR: 0.5-0.9) and osmolality 77 mOsm (IQR: 57-89).

Intra-individual variability

Short chain fatty acids, such as hexanoic- and pentanoic acids were most consistent across the 3-week period, with the coefficient of variation (CV) demonstrating acceptable levels below 20%. Other compounds, acetic-, butyric- and propanoic acids showed higher variability with CV reaching 50% in certain individuals. Two patterns of interest were identified: (i) consistency for all 5 samples across the 3 weeks and (ii) variability of one sample only.

Two participants (ID: S6 and S7) demonstrated a consistent concentration of fatty acids across 3 weeks. Butyric acid ranged from 2.83 to 3.21 ppbv (CV 5%) and propanoic acid from 5.31 to 7.99 ppbv (CV 17%). The variability was further reduced when assessing the week-to-week variability with similar concentrations of propanoic acid detected: 6.4, 5.7 and 6.0 ppbv in weeks one to three respectively. A similar trend was observed for aldehydes with an acceptable level of concentrations between days and weeks of saliva sampled. In S7, butanal ranged from 3.32 to 4.42 ppbv (CV 12%), and pentanal from 8.26 to 14.0 ppbv (CV 19%). The second trend identified was owing to one sample demonstrating much higher VOC concentrations. For example (ID: S2), butyric acid detected across all 5 samples were at levels 10.4, 4.9, 5.1, 3.9 and 3.0 ppbv. The first sample measured an osmolality of 157 mOsm compared to the remainder of the samples measuring between 65 – 89 mOsm. A higher osmolality was associated with a higher VOC concentration. This was also true for aldehydes, in particular pentanal and propanal. Pentanal levels reduced from 36 ppbv in the first sample to an average of 13 ppbv in the following 4 samples. This trend was also observed in other participants. With the exception of specific compounds such as acetaldehyde and propanal, there was an acceptable variation in levels of selected VOCs over the five assessed time points. All values are presented in the Appendix 13.4 (Figure 22).

Inter-individual variability

Inter-individual variability, assessing daily and weekly variation over 3 weeks, was minimal for all target compounds. Butyric acid had a CV 6.3% (concentrations of 4.2, 4.1, 3.6, 3.8, 3.9 ppbv; $p=0.28$) suggesting adequate reproducibility and limited influence from external factors. Butanal and pentanal also showed a narrow range of variation with a CV of 12.5% ($p=0.05$) and 10.1% ($p=0.3$) respectively. Five of the 15 compounds were statistically different, although CV% was deemed acceptable. For

example, acetic acid CV 9.6%, $p=0.03$, propanoic acid CV 15.5%, $p<0.001$ and heptanal CV 12.9%, $p<0.02$ (Table 19, Figure 23).

Table 19. Longitudinal variation of the concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva, as a cohort. Samples: 1 = day 1, 2 = day 2, 3 = day 3, 4 = day 10 4, 5 = day 17. Coefficient of variation percentage, CV%.

Compound	Sample concentration (ppbv)					CV%	P value
	Day 1	Day 2	Day 3	Day 10	Day 17		
Acetone	179	145	184	196	246	19	0.06
Short chain fatty acids							
Acetic acid	8.2	6.7	6.5	6.9	6.7	9.6	0.03
Butanoic acid	4.2	4.1	3.6	3.8	3.9	6.3	0.28
Hexanoic acid	1.4	1.3	1.0	1.2	1.0	15	0.15
Pentanoic acid	4.3	3.7	3.7	3.5	3.8	7.9	0.12
Propanoic acid	5.3	4.2	3.9	3.6	4.2	16	<0.001
Aldehydes							
Acetaldehyde	153	175	207	135	210	19	0.049
Butanal	3.9	3.1	3.6	3.1	2.8	13	0.05
Decanal	1.6	1.5	1.8	1.8	1.5	8.1	0.71
Heptanal	2.4	2.3	2.6	2.5	1.8	13	0.02
Hexanal	6.7	4.8	5.6	4.3	4.5	19	0.4
Nonanal	3.2	2.9	3.4	3.1	2.9	6.7	0.03
Octanal	2.8	2.2	2.6	2.3	2.2	11	0.69
Pentanal	14.8	13.5	12.0	11.5	12.6	10	0.3
Propanal	28.2	20.0	28.5	19.9	28.3	18.4	0.69

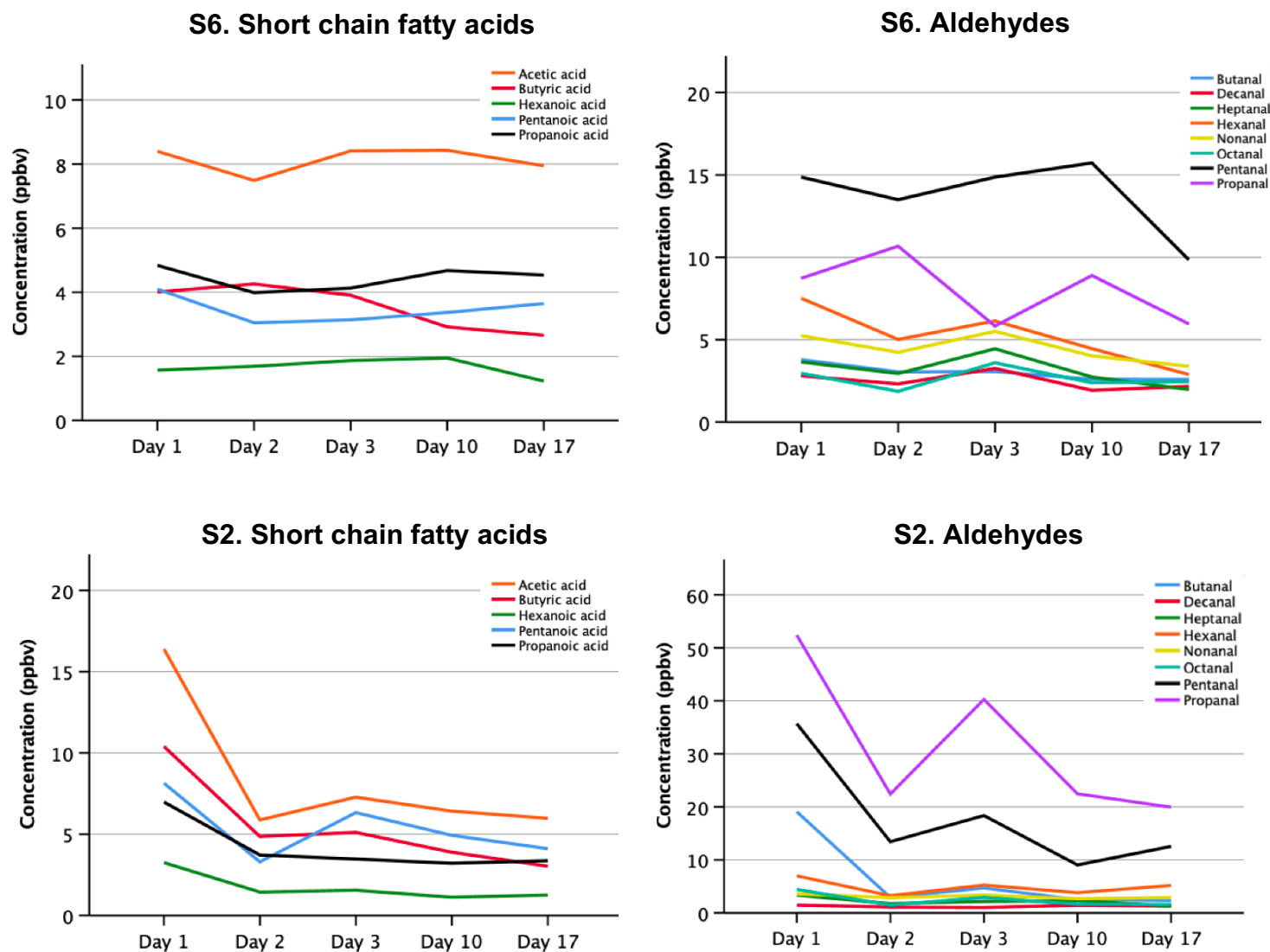


Figure 22. Longitudinal variation of target short chain fatty acids and aldehydes in human saliva. S6 = participant 6, S2 = participant 2. There is an acceptable level of variation of volatile compounds over the five time points assessed.

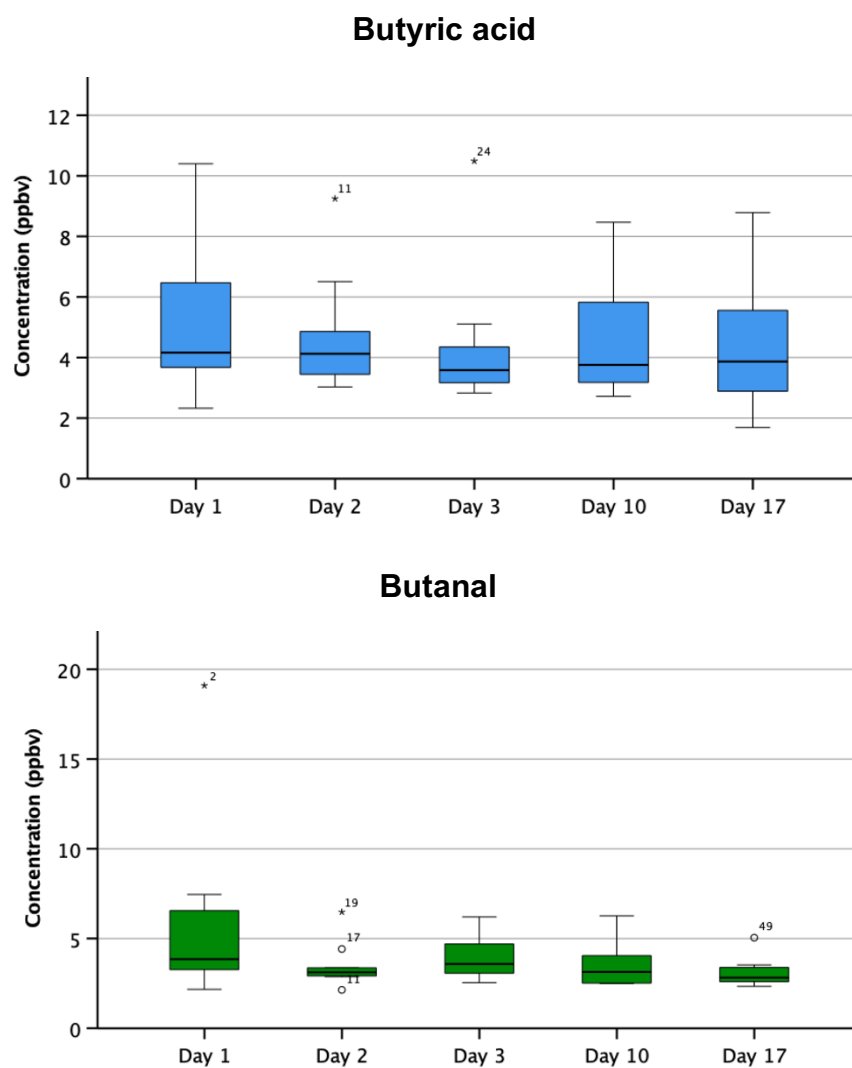


Figure 23. Overall longitudinal variation of the concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva, as a cohort. There is an acceptable variation in concentration of select compounds across the cohort.

7.2.5 Discussion

The primary finding of this study is that salivary VOCs have an acceptable level of day-to-day and week-to-week variation, with the exception of select compounds. To the best of my knowledge, this is the first study assessing longitudinal VOC variation in saliva. In this study, acetaldehyde and propanal showed wide variation over the 3 weeks with a high coefficient of variation greater than 50%. This suggests that certain compound levels may be labile either due to their volatility or systemic factors. However, the majority of compounds were stable with a comparable median. It is also important to note the range of VOCs as this may be different in a control group compared to a diseased group. A narrow variation of concentrations may not be as important as the differences in levels between both control and disease groups. Distinguishingly higher levels of certain VOCs can mean that a wider variation of 'normal' VOCs is acceptable. This study confirms that assessment of salivary VOCs is a feasible approach in a clinical diagnostic setting.

7.2.6 Conclusion

There is minimal inter-individual longitudinal variation of salivary VOCs with an acceptable coefficient of variation. However, select compounds can vary significantly within an individual over time. It is important to establish a normal range of VOCs in a healthy population to ascertain clinically significant VOCs.

8. CHAPTER 8: ASSESSMENT OF SALIVARY HEADSPACE VOLATILE ORGANIC COMPOUNDS IN OESOPHAGOGASTRIC CANCER

8.1 ASSESSMENT OF SALIVARY VOCs FOR THE DETECTION OF OESOPHAGOGASTRIC CANCER USING TARGETED SIFT-MS

8.1.1 Introduction

There is paucity of information about the use of salivary VOCs in disease diagnostics. Recent interest in non-invasive investigations has provided a platform for the potential use of saliva in early gastrointestinal cancer detection. In particular, benign upper gastrointestinal symptoms overlap with early oesophagogastric cancer and therefore there is a clinical unmet need for a triage test to identify high-risk patients. Our group have proposed a VOC profile of exhaled breath for the detection of oesophagogastric cancer, with a sensitivity of 80%. [32] A contribution of VOCs emanating from the oral cavity is now recognised; as detailed in Section 3 Chapter 6.1. Recent studies have assessed the use of salivary VOCs in localised oral disease, and systemically for the detection of gastrointestinal cancers. [223, 224] A VOC profile for detection of stomach and colorectal cancer with sensitivities of 80% and 92.3% respectively have been demonstrated. [224]

8.1.2 Aim

This study aims to identify a VOC profile specific to oesophagogastric cancer patients compared to healthy controls.

8.1.3 Methods

Participant selection Patients diagnosed with oesophagogastric cancer (oesophageal/GOJ cancer n=25, gastric cancer n=25) and healthy controls (n=25) were recruited from September 2020 to August 2021. Inclusion criteria included: biopsy proven adenocarcinoma of the oesophagogastric tract, on the curative pathway, aged over 18 years of age, able to provide written informed consent. Exclusion criteria included: known oral disease, known severe systemic disease including concurrent cancer, use of immunosuppressive medications. All patients were recruited on the morning of their surgery/endoscopic procedure and therefore had been fasted for a minimum of 6 hours. Normal oral hygiene routines were maintained with a minimum of one hour before saliva sampling. All participants provided written informed consent. Ethical approval was obtained via the Imperial College Healthcare Tissue Bank; REC Wales approval: 17/WA/0161 (R20059).

Sample processing A minimum of 500µL of passive drool saliva samples were collected into a sterile container. Saliva was immediately aliquoted into 1.8mL cryovials and stored at -80°C until analysis. Prior to analysis all saliva samples were thawed at room temperature. 500µL aliquots were transferred

into 20mL glass headspace vials secured with 18mm metal screw-tops and PTFE septa. All samples were incubated at 37°C for 30 minutes in a water bath to equilibrate the VOCs between the aqueous and gaseous phases.

Mass spectrometry analysis SIFT-MS Direct sampling was performed using the SIFT-MS with a targeted method for volatile short chain fatty acids (C2-C6), previously established by the group. The septa were pierced with a sterile blunt needle (18G, 1 ½ (1.2mm x 40mm)) with an analytical time frame of 30 seconds and an inlet flow rate of 25 mL/min, extracting 12.5 mL of headspace.

Statistical analysis Mann Whitney U test was performed to calculate statistical significances between cancer and non-cancer groups (IBM SPSS Statistics version 25 (Armonk, NY)).

8.1.4 Results

Thirty patients (gastric adenocarcinoma n=10, oesophageal adenocarcinoma n=10, healthy controls n=10) were recruited between September 2020 to August 2021. Baseline characteristics of the patients are detailed in Table 20.

Table 20. Baseline characteristics of patients between the three groups.

	Gastric adenocarcinoma	Oesophageal adenocarcinoma	Controls
Cases, n (%)	10 (33.3)	10 (33.3)	10 (33.3)
Age, y (IQR)	73 (56-80)	71 (66-75)	52 (47-60)
Gender (M:F)	8:2	10:0	3:7
Ethnicity			
White	7	10	6
Asian	1	0	3
Arab	1	0	1
Black	1	0	0
Clinical staging, n			-
T1	2	1	
T2	2	2	
T3	2	6	
T4	3	1	
N0	5	2	
N1	4	6	
N2	1	2	
N3	0	0	
M0	10	10	
M1	0	0	
Tumour Differentiation, n			-
Poor	5	3	
Moderate	5	4	
Well	0	0	
Unknown	0	3	
Controls: diagnosis			
Gallstone disease	-	-	4
Reflux			4
Hernia			2

Butyric acid was detected in significantly higher concentrations in the headspace of saliva of cancer patients. Levels of 8.8 ppbv (IQR: 7.2-11.6 ppbv) were identified in healthy controls, compared to 2-fold (15.6 ppbv (IQR: 10.9-21.1), $p=0.014$) and 3-fold (27.3 ppbv (IQR: 13.1-66.4), $p=0.002$) higher levels in oesophageal and gastric cancer, respectively. Propanoic acid was significantly lower in gastric cancer (4.0 ppbv (IQR: 3.5-5.3), $p<0.001$) and oesophageal cancer (6.2 ppbv (IQR: 5.6-7.0), $p=0.003$) compared to healthy controls (8.7 ppbv (IQR: 7.9-9.6)). Hexanoic acid was also reduced in gastric cancer compared to healthy controls. No differences were observed with the remainder of the fatty acids and aldehydes assessed (Table 21, Figure 24).

Table 21. Median concentrations (ppbv) of target short chain fatty acids and aldehydes in saliva detected using the SIFT-MS (VoiceUltra200).

Compound	Median concentration, ppbv (IQR)			P value	P value
	Controls	Oesophageal adenocarcinoma		Gastric adenocarcinoma	
Acetone	371 (332-462)	437 (286-576)	0.97	435 (346-647)	0.62
Short chain fatty acids					
Acetic acid	10.7 (8.9-20.9)	10.8 (9.0-11.7)	0.73	8.2 (6.2-9.7)	0.09
Butyric acid	8.8 (7.2-11.6)	15.6 (10.9-21.1)	0.014	27.3 (13.1-66.4)	0.002
Hexanoic acid	1.7 (1.3-2.4)	1.7 (1.0-1.8)	0.19	1.3 (1.1-1.4)	0.04
Pentanoic acid	8.5 (6.7-10.5)	11.8 (9.0-14.4)	0.06	7.1 (5.7-12.1)	0.67
Propanoic acid	8.7 (7.9-9.6)	6.2 (5.6-7.0)	0.003	4.0 (3.5-5.3)	<0.001
Aldehydes					
Acetaldehyde	196 (102-237)	245 (160-341)	0.24	168 (160-212)	0.97
Butanal	3.7 (2.8-4.4)	5.4 (3.6-6.9)	0.17	3.3 (2.6-4.6)	0.88
Decanal	1.1 (1.0-1.2)	1.3 (1.1-1.4)	0.21	1.0 (0.9-1.1)	0.19
Heptanal	4.5 (2.8-5.9)	5.7 (4.2-6.7)	0.43	4.1 (2.1-6.2)	0.67
Hexanal	4.9 (3.1-6.4)	6.6 (4.3-8.8)	0.24	5.2 (3.8-7.1)	0.70
Nonanal	2.8 (2.4-2.9)	2.4 (2.4-2.8)	0.70	2.4 (1.7-2.7)	0.19
Octanal	2.2 (2.0-2.7)	2.7 (2.2-3.0)	0.45	2.2 (1.4-3.1)	0.91
Pentanal	4.4 (3.0-5.3)	5.7 (3.7-6.7)	0.16	3.9 (3.2-4.8)	0.97
Propanal	42.2 (24.9-60.3)	45.0 (21.4-67.5)	0.73	19.3 (14.7-64.9)	0.52

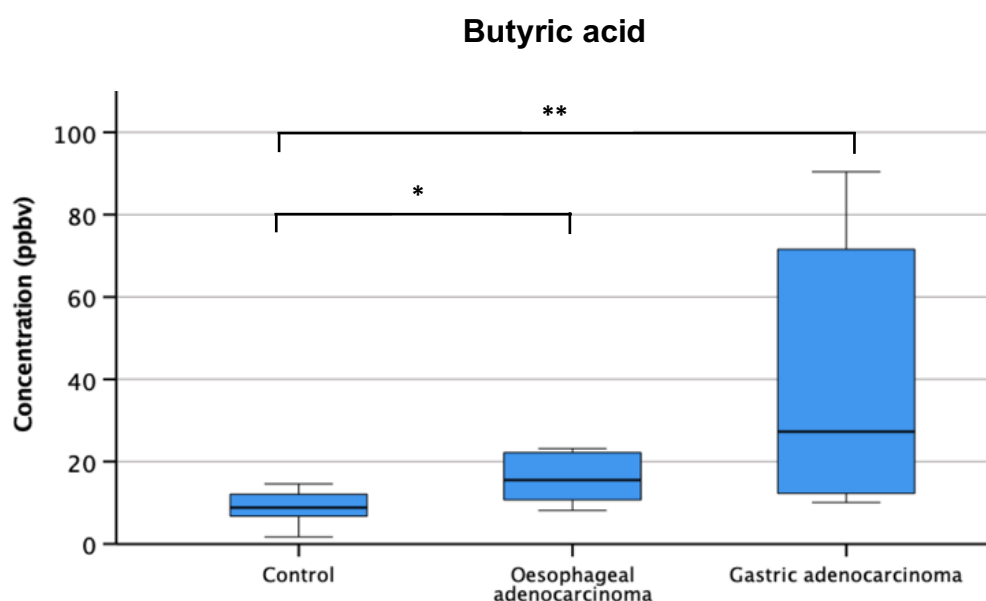


Figure 24. Box plot showing levels of butyric acid detected within whole saliva are significantly higher in the gastric and oesophageal adenocarcinoma groups compared to healthy controls. * $p<0.05$, ** $p<0.01$.

8.1.5 Discussion

Saliva presents a potentially non-invasive method of detecting disease. The primary finding of this study is increased levels of butyric acid from whole saliva in oesophagogastric cancer, compared to healthy controls. Butyric acid is an important potential biomarker in oesophagogastric cancer evidenced by work in our group demonstrating elevated butyric acid levels in (i) breath and (ii) direct headspace above the tumour. In addition, this study raises the possibility that the oral cavity may contribute to butyric acid levels initially thought to originate wholly from exhaled breath.

Butyric acid has been linked to poor oral hygiene and periodontal disease, predominantly arising from microbial metabolism within the mouth. Sources identified have been plaque, tongue coating and food debris. To eliminate influence from external factors such as food/drink consumption and oral hygiene products, all participants were fasted for >6 hours, and had cleaned their mouth at least one hour prior to sampling. This was to mitigate VOCs from exogenous sources and ensure any VOCs detected were wholly from a 'clean' oral cavity. It is important to recognise that recent research has revealed cancer-specific microbial profiles within the oral cavity of cancer patients. Therefore, bacterial degradation here may produce a distinctive VOC profile.

There were several limitations noted with sample collection for this study. Several patients reported an extremely dry mouth having fasted overnight for their scheduled surgery/endoscopic procedure and had difficulty producing saliva. These patients were given approximately 5 mL of water to stimulate some saliva and proved to be successful. Secondly, the saliva composition was different between samples. Some contained more mucus and was more concentrated, and others a uniform liquid.

8.1.6 Conclusion

Targeted saliva headspace VOC analysis has shown a significantly higher concentration of butyric acid specific to oesophageal and gastric cancers. There may be a contribution from the oral cavity to the original data proposing butyric acid as a biomarker in exhaled breath in oesophagogastric cancers.

8.2 ASSESSMENT OF SALIVARY VOCs FOR THE DETECTION OF OESOPHAGOGASTRIC CANCER USING UNTARGETED 1D GC-MS

8.2.1 Introduction

Further to the targeted volatile short chain fatty acid saliva study, it was important to investigate other VOCs which may be important to the early detection of oesophagogastric cancer. The earlier study demonstrated a significant increase of butyric acid in both gastric adenocarcinoma and oesophageal adenocarcinoma groups, compared to healthy controls. To investigate the wider group of VOCs, an untargeted discovery study was designed, with the use of the Centri® GC-MS, an automated system for high throughput headspace analysis.

8.2.2 Aims

This study aims to identify a VOCs specific to oesophagogastric cancer patients compared to healthy controls.

8.2.3 Methods

Whole saliva was obtained from patients diagnosed with oesophageal adenocarcinoma (n=21), gastric adenocarcinoma (n=22) and healthy controls (n=22). Samples were thawed on ice and 500µL aliquoted into 20mL glass headspace vials. Each sample was spiked with an internal standard mix containing 6 deuterated compounds: benzaldehyde-d6, butyric acid-d8, toluene-d8, acetaphenone-d8, 2-propanol-d8, octane-d18. 10µL of a 5ppm solution was spiked into a 500µL sample to achieve a 100ppb detection rate. Two quality indicators were used after every 11 samples; (i) pooled samples consisting of 5 healthy control saliva samples, also spiked with an internal standard mix (ii) a standard QC mix of 10µL into 500µL of water plus internal standard. The standard QC mix consisted of 10 compounds: pentanoic acid, butyric acid, 2-hexanone, 2-cotanone, decanal, hexanal, 1-nonanol, butanol, dodecane, tridecane. Additional empty vials, vials containing water, and vials containing water plus an internal standard were used to assess the integrity of the samples and detect any instrument shift.

The Centri® (GERTSEL) and Agilent 6890N gas chromatograph was connected to a BenchTOF2 (Agilent Technologies) and fitted with a column allowing detection for a wider range of mass compounds (Column: Restek Rxi-624Sil MS – 20m x 0.18mmID x 1.0µm film thickness). The minimum nitrogen carrier pressure was set at 5 psi. 20mL glass headspace vials containing 500µL of samples were incubated and agitated at 100°C for 30 minutes at 300rpm. The instrument was unable to reach a physiological temperature of 37°C because of the adjacent high temperature of 100°C of the injection inlet. The Centri® syringe extracted 5mL of headspace volume with a syringe temperature of 100°C. Samples were enriched onto the cold trap with 2 full injections of 5mL each. A 1mL nitrogen purge of

the syringe at a pressure of 0.2 bar was set between each sample. The sample was injected at 200°C over 3 minutes onto the headspace trap (300°C, flow rate 50mL/min). The column flow rate was 1.5mL/min (30 psi). The GC oven temperature program for desorption was 40°C (hold for 1 minute), then a 10°C/min ramp to 250°C (final hold of 5 minutes). The transfer line temperature was 280°C. The BenchTOF2 was optimised for positive electron ionisation (EI) standard AO mode (70 eV) with a mass range of 30-600 atomic mass units. Standard quality checks in keeping with industrial standards were performed for optimisation of human saliva samples.

Data was generated using ChromSpace software with a minimum ion count of 500 to exclude background noise and filtering compounds with a reverse match factor greater than 800. Mann Whitney U test was performed to calculate statistical significances between cancer and non-cancer groups (IBM SPSS Statistics version 25 (Armonk, NY)).

8.2.4 Results

Sixty-five patients (gastric adenocarcinoma n=22, oesophageal adenocarcinoma n=21, healthy controls n=22) were recruited between December 2020 to September 2021. Baseline characteristics of the patients are detailed in Table 22.

Table 22. Baseline characteristics of patients between the three groups.

	Gastric adenocarcinoma	Oesophageal adenocarcinoma	Controls
Cases, n (%)	22 (34)	21 (32)	22 (34)
Age, y (IQR)	74 (58-79)	69 (64-75)	63 (57-69)
Gender (M:F)	10:12	17:4	9:13
Ethnicity			
White	14	18	10
Asian	3	3	2
Arab	0	0	7
Black	5	0	3
Clinical staging, n			-
T1	1	0	
T2	5	7	
T3	10	10	
T4	5	4	
N0	7	6	
N1	8	10	
N2	5	2	
N3	2	3	
M0	22	21	
M1	0	0	
Tumour Differentiation, n			-
Poor	11	6	
Moderate	8	10	
Well	0	0	
Unknown	3	5	
Controls: diagnosis			
Gallstone disease	-	-	3
Reflux			4
No abnormalities detected			15

No significant shifts in the instrument performance were noted. All 5 QC samples spiked with internal standards were detected at comparable concentrations with an acceptable CV% of less than 30%, except one: 2-propanol-d8 CV 24%, acetaphenone-d8 CV 4%, octane-d18 CV 11%, toluene-d8 CV 7%, benzaldehyde CV 61%.

A total of 69 compounds were identified across all samples. 7 compounds were statistically significant between cancer and non-cancer groups. Propanoic acid-ethyl ester was found in significantly higher abundances within the saliva of gastric adenocarcinoma patients (median peak area = 2,779,262; $p=0.04$) compared with healthy controls (median peak area = 3,578,550). (Figure 25 and 26). Other compounds included: 2-propenyl-benzene ($p=0.03$, reduced in gastric cancer), 1,1-dimethoxy-2-phenylpropane ($p=0.004$, reduced in gastric cancer), 2-chloroethyl methyl ether ($p=0.004$ elevated in oesophageal cancer), branched benzeneethanamine ($p=0.006$ and $p=0.04$ elevated in gastric cancer and oesophageal cancer respectively), o-xylene ($p=0.03$ elevated in gastric cancer), bis(trifluoromethyl) trioxide ($p=0.02$ elevated in gastric cancer).

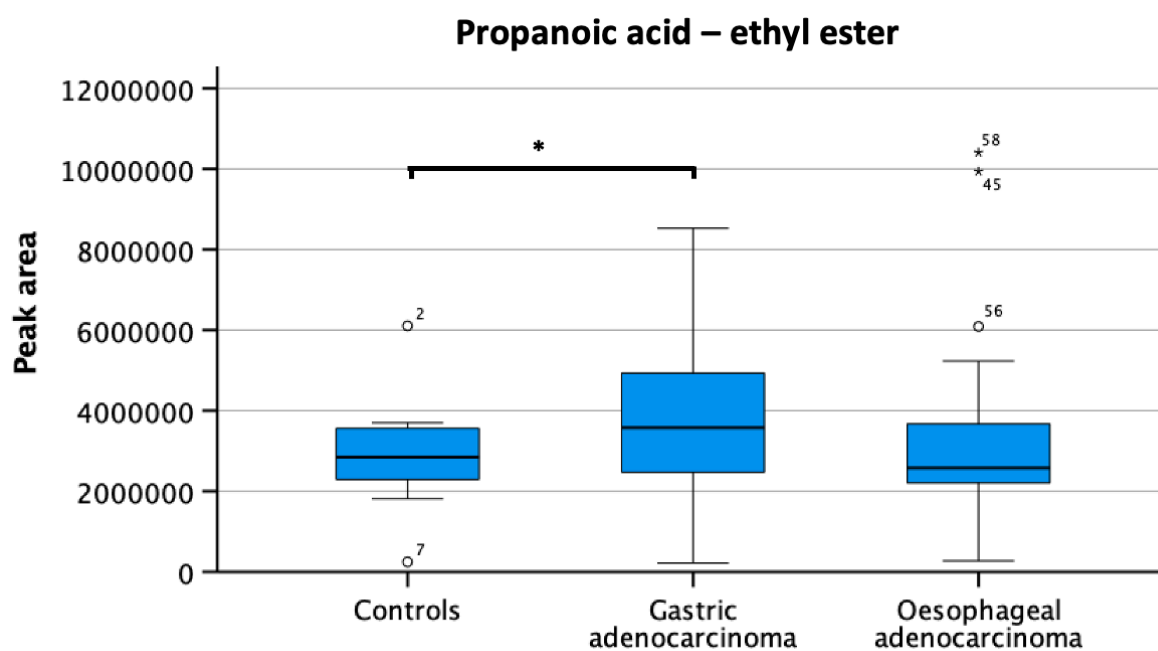


Figure 25. Box plot demonstrating the peak area of propanoic acid – ethyl ester in the headspace of saliva of oesophagogastric cancer and healthy control patients. Statistical significance was demonstrated with elevated concentrations of propanoic acid – ethyl ester in gastric cancer ($p=0.04$) compared to healthy controls.

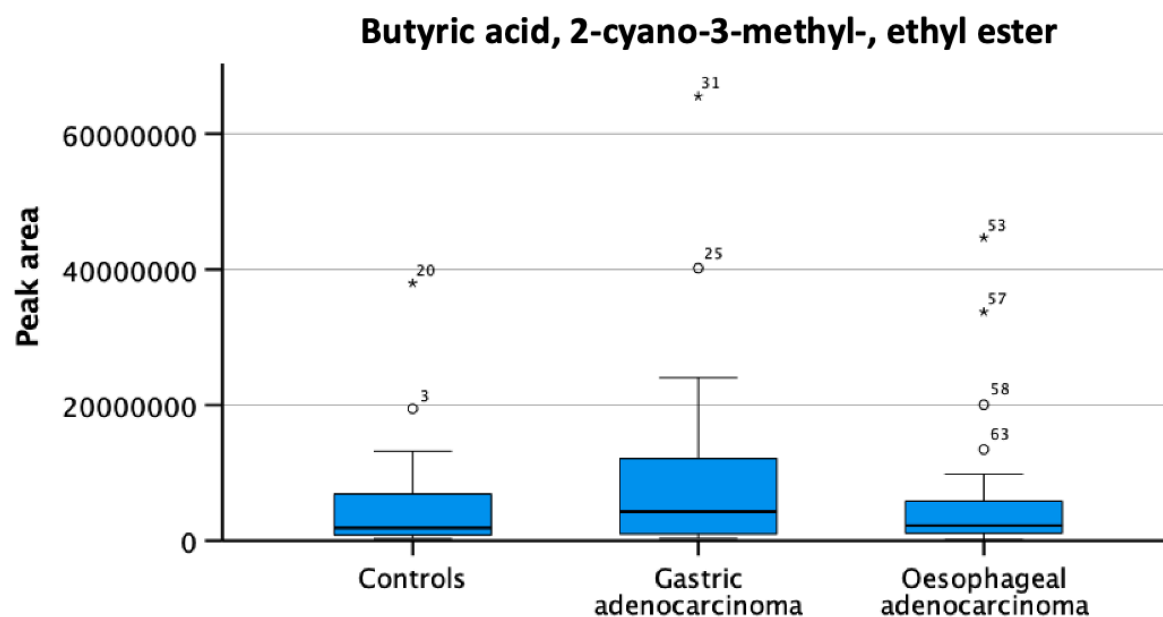


Figure 26. Box plot demonstrating the peak area of a branched butyric acid (-2-cyano-3-methyl-, ethyl ester) in the headspace of saliva of oesophagogastric cancer and healthy control patients. Median peak area calculation of gastric adenocarcinoma (median peak area = 4,292,933) is markedly higher than healthy controls (median peak area = 1,919,955), albeit not significant.

8.2.5 Discussion

The primary significant finding of this study is the elevation of propanoic acid – ethyl ester in the headspace of saliva from gastric cancer patients compared to healthy controls. Other significant branched compounds were also identified, however propanoic acid – ethyl ester is of particular interest. Propanoic acid – ethyl ester, also known as ethyl propionate, is a volatile liquid which is synthesised by the esterification of alcohol and propanoic acid.

Compounds of interest, those related to butyric acid, were identified in higher abundances in gastric cancer (median peak area = 4,292,933) and oesophageal cancer (median peak area = 2,238,590) compared to healthy controls (median peak area = 1,919,955). However, they did not reach statistical significance. With higher recruitment numbers, it is possible significance may be reached. This finding is in keeping with earlier results demonstrating significantly higher concentrations of butyric acid in the gastric cancer group using the SIFT-MS platform. The differences observed are likely multi-factorial, from sample preparation (higher temperatures for Centri 1D GC-MS analysis directly affecting volatility of lower molecular weight compounds) to analytical framework with shorter direct sampling technique using the SIFT-MS. This may translate to reduced likelihood of sample loss or sample degradation.

A study by Bel'skaya *et al* demonstrated high concentrations of ethanol, methanol, and acetaldehyde in patients with gastric cancer compared to controls, albeit not statistically significant. A classification tree was constructed and 4 VOCs in combination were selected to build a diagnostic VOC model. Acetaldehyde, acetone, 2-propanol and ethanol were selected to give an AUC of 0.839 for gastric cancer.

In analytical chemistry, it is important to recognise shifts in instrument performance over time to account for differences in peak areas of volatile compounds detected. Pooled saliva samples spiked with internal standards were used as a quality indicator after every 11 samples of the total 65. An acceptable coefficient of variation demonstrated over the 3 sampling days meant the instrument performance was stable. There is unlikely to be a significant shift over a shorter sampling period, however it remains good practice to assess for this. If significant changes were observed, peak area correction would have been performed against the internal standards used for different chemical classes.

8.2.6 Conclusion

This study reports the first use of an automated sampling system in our group; the Centri® autosampler. Optimisation studies were performed to ensure VOCs could be detected after headspace extraction and injection into the GC system. As this was the first time an untargeted discovery approach was used for salivary VOCs, a profile of significant potential biomarkers was identified, with particular interest in propanoic acid related compounds.

SECTION 4

AUGMENTATION OF VOLATILE ORGANIC COMPOUNDS IN OESOPHAGOGASTRIC CANCER

9. CHAPTER 9: *IN VITRO* AUGMENTATION OF VOC FROM OESOPHAGOGASTRIC CANCER BACTERIA CULTURES

9.1 INTRODUCTION

Augmentation of short chain fatty acids by manipulating bacterial metabolism is a well-established practice in industry. Optimal fermentation parameters within a large-scale bioreactor allows maximum yield of predominantly butyric-, acetic- and propanoic acids. Based on this and the VOC breath profile of oesophagogastric cancer patients (including butyric acid and pentanoic acid), a study was designed to simulate this augmentation in an *in vitro* environment. This study is intended to isolate the role of the cancer associated microbiome and its potential contribution to an augmented exhaled breath test.

9.2 AIM

The primary aim is to identify headspace VOCs related to (i) cancer-specific bacteria alone and (ii) cancer-specific bacteria exposed to nutritional substrates. This will be achieved by establishing cell growth curves to identify time points for optimal metabolic activity when VOCs can be extracted.

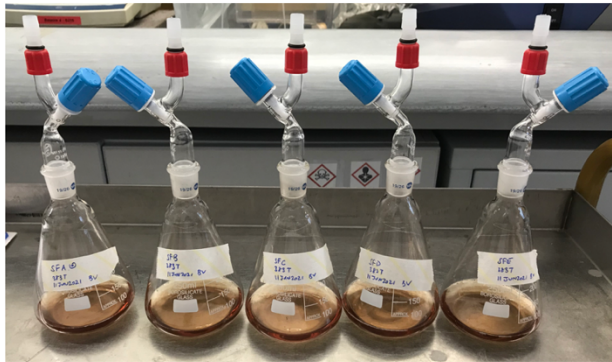
9.3 METHODS

Liquid cultures were established to monitor cell growth and extract headspace VOCs. Fastidious anaerobic broth (Neogen CM0199A) and Nutrient Broth (Oxoid CM001) was used for anaerobic and aerobic cultures, respectively. 50µL of inoculum (bacteria cultivated from oesophagogastric tissue and saliva is described in Section 2, Chapter 5.1.3) was added to 50mL of media in a 250mL round bottom flask fitted with a stopcock adapter for anaerobic cultures, and a 250mL baffled flask for aerobic cultures. Anaerobic cultures were sparged with pure nitrogen (99.99%) into the inoculated media and headspace at a flow of 0.5L per minute for 1 minute through a 0.2µm filter. Five starter flasks were set-up for each experiment to produce high density of bacteria growth to allow sacrificial flasks for the planned augmentation experiments (Figure 27). Shake flasks were incubated at 37°C at 250 rpm on the Aquila Biosystem for real-time cell growth quantification (Figure 28). Aquila Biosystems uses the principle of backscatter of light. Optical density and headspace VOC measurements were taken pre-culture and post-culture for two representative flasks. Satisfactory cell growth was determined by optical density measurements, macroscopic assessment of the liquid colour (density, colour, homogeneity) with review of the cell growth assays. Contents of each flask were transferred to a 50mL falcon tube and centrifuged at 5000g at 4°C for 15 minutes. The supernatant was discarded, and the cell pellet resuspended in 50mL of sterile PBS (pH 5.0 for cultures obtained from the oesophagogastric tract). The suspensions were incubated for a period of 1 hour at 37°C at 250 rpm to “starve” the cells and deplete all potential carbon sources from the culture medium. They were centrifuged at 3500g at 4°C for 10 minutes, and the supernatant discarded. Cells were resuspended in sterile PBS and all 5 bacterial

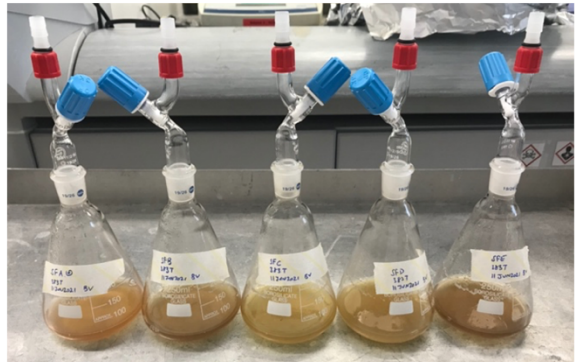
suspensions were combined to mix into a single flask. 50mL of the suspension was aliquoted into 5 different 250mL flasks. 1g of filter-sterilised glucose was added to 4 of the 5 flasks, to a concentration of 0.1M. Anaerobic samples were sparged with pure nitrogen (99.99%) at a flow of 0.5L per minute for 1 minute through a 0.2µm filter. A stopcock adapter was used to seal the system. Flasks were incubated at 37°C at 250 rpm for 10-, 30-, 60-, and 120 minutes. The glucose-free flask was incubated to allow the headspace to reach equilibrium. At each time interval, TD tubes were connected to a stopcock adapter with C-flex tubing, the adapter tap opened and 200mL of headspace extracted at a flow rate of 200mL/minute using a pocket pump. Each flask was discarded after headspace extraction so each time point would be truly reflective of the VOC concentration (Figure 29).

Anaerobic cultures

Pre-culture

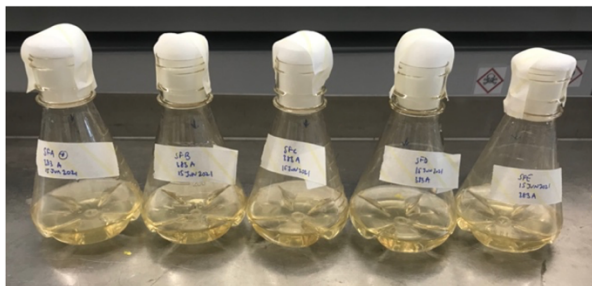


Post-culture



Aerobic cultures

Pre-culture



Post-culture



Glucose Augmentation set-up



Figure 27. 250mL flasks for culture growth and glucose augmentation set-up. Stopcock adapters: Tap (blue) to release the seal to allow headspace extraction via the outlet (red). Round bottom flasks used for anaerobic cultures, and baffled flasks for aerobic cultures.

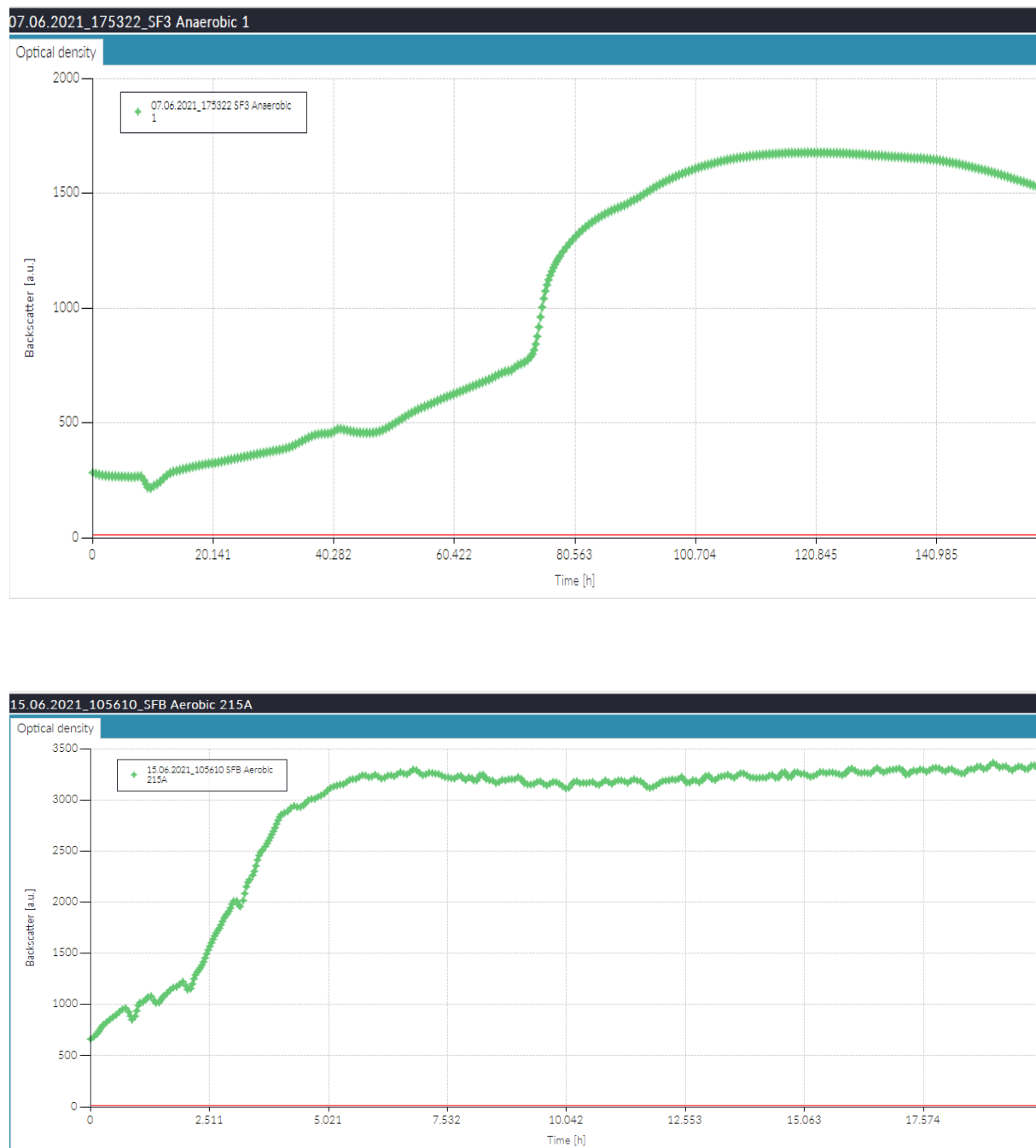


Figure 28. Cell growth curves of mixed and pure bacteria cell cultures in liquid media. Growth assays depict 4 phases of the cell cycle: (i) lag phase (ii) log/exponential phase (iii) stationary phase and (iv) death phase, for anaerobic and aerobic mixed cultures.

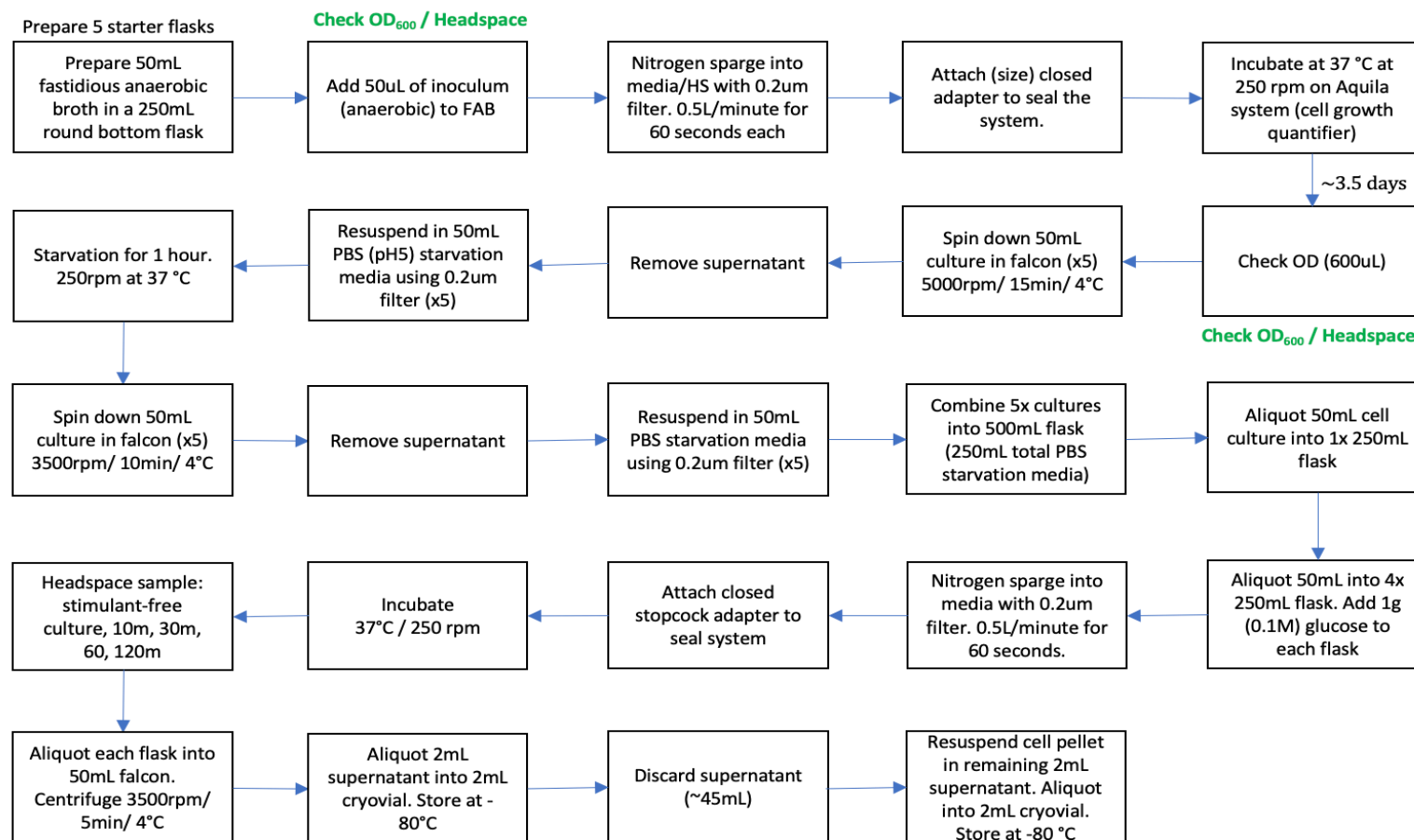


Figure 299. Biotransformation workflow describing the VOC augmentation from liquid bacteria cell cultures.

Following analysis, the bacterial suspension was transferred into a 50mL falcon tube and centrifuged at 3500g for 10 minutes. 2mL of the supernatant was transferred into a 2mL cryovial. 46mL was discarded, and the cell pellet was resuspended in the remaining 2mL and transferred to a 2mL cryovial. Samples were stored at -80°C until further processing.

Analytical strategy Bacterial DNA was extracted from the cell pellets as per the FastDNA Spin® kits (MP Biomedicals), as described in Chapter 5.1. Mixed bacterial communities from each time interval were sent for Shotgun sequencing, and one pure culture for 16S gene Sanger sequencing (Macrogen®). **Metagenomic Sequencing** Mixed bacteria culture samples underwent Shotgun metagenomic sequencing using the Illumina Nextera XT kit and Illumina Novaseq 150bp PE (2Gb data per sample). Pure cultures underwent Illumina 16S rRNA gene (V3-V4) amplification and Illumina Miseq sequencing 300 PE (approximately 100,000 reads per sample; Macrogen®, South Korea).

Characterisation of metagenomes derived from mixed bacterial cultivation work. The quality of the raw sequence data (paired-end 150 bp) was assessed using fastQC v0.11.9; the data did not require trimming. An average of 2.91 Gb ($\pm$ 0.39 Gb) of sequence data were generated for each sample. To determine taxonomic coverage within samples, paired-end read data were run through Kraken2 2.0.8-beta [186] using the pre-compiled Kraken2 GTDB_r89_54k index (downloaded on 3 May 2020) available from <https://bridges.monash.edu/ndownloader/files/16378439> [187], with data summarized at different taxonomic levels (phylum to species). This gives a broad overview of microbes present within a sample. To generate sample-specific data, metagenome-assembled genomes (MAGs) needed to be assembled. The sequence data were assembled using MEGAHIT v1.2.9 [188]. Raw sequence reads were mapped against the assembled contigs and MetaBAT2 v2.12.1 was used to bin the data [189]. A total of 727 bins were created across all samples. MAGpurify v2.1.2 (database v1.0) was used to clean up the bins (i.e. the MAGs) [190]. The taxonomic affiliations of the microbes represented in the MAGs were determined using ANI analysis with GTDB-Tk v1.3.0 (release 95).[191] Assessments of completeness and contamination of the MAGs were made using CheckM v1.13.0; MAGs were annotated using Prokka v1.14.0 and summary statistics were determined for all. Overall quality of MAGs was assessed according to Bowers *et al.*: those with ≥ 90 % completeness and ≤ 5 % contamination were considered high-quality draft genomes; those with ≥ 50 % completeness and ≤ 10 % contamination were considered medium-quality draft genomes; all other MAGs were considered low quality and were not examined further.[192] A phylogenetic tree incorporating the medium- and high-quality MAGs was generated using PhyloPhlAn v0.99 and visualized using iTOL v6.3.[193] The primary metabolites encoded within the medium- and high-quality MAGs were predicted using gutSMASH (Pfam_A.hmm, 2 May 2020).[194]

Characterization of pure isolates recovered from bacteria cultivation work. Trimmed raw sequence data (microbesNG) were checked using fastQC v0.11.9. Further editing was not required. Sequence data were assembled using SPAdes v3.11.1.[195] Contigs <200 nucleotides in length were

removed from assemblies using reformat.sh in BBMap. Prokka v1.14.0 was used to predict coding sequences encoded within the draft genomes.[196] CheckM v1.1.3 was used to determine the completeness and contamination of the genomes.[197] Tentative identities for the assembled genomes were determined using PubMLST Species ID.[198] Taxonomy and metadata files for release 202 of the Genome Taxonomy Database were downloaded, and used to identify type species sequences for taxa most closely related to the three isolates sequenced by microbesNG.[199] The relevant 113 representative high-quality genomes were downloaded from GenBank and used to create a database against which the three newly generated genome sequences were compared: (i) for average nucleotide identity (ANI) analysis using fastANI [200], then (ii) phylogenetic analysis using PhyloPhlAn v0.99.[201]

The supernatant was sent to the National Phenome Centre (Imperial College London) for metabonomic analysis. Thermal desorption tubes were analysed on the two-dimensional gas chromatography-mass spectrometer (GCxGC-MS, Markes®-TD100-XR - Agilent 6890 - BenchTOF Select, SepSolve, UK). VOC's were separated upon a MegaWax HT column in the first dimension (20m x 0.18mm x 0.18µm) and a VF-200ms column in the second dimension (5m x 0.25mm x 0.10µm). The system was programmed at 0.7mL/min constant Helium carrier. Oven temperature was set at 55°C initially for 1 minute, ramp to 240°C at 4°C/minute with a 2 minute hold. The mass analyser was set to acquire over the range of 33 to 600 m/z. The ion source and transfer line were maintained at 260 °C. Compound identity was confirmed by spectral analysis of the primary ions and its likely presence in bacteria.

9.4 RESULTS

All mixed bacteria cultures and pure cultures were cultivated to their exponential growth phase guided by cell growth curves using the Aquila Biosystem, and optical density measurements. Some cultures were more easily cultured than others. For example, *Escherichia coli* grew more rapidly as evidenced by an optical density of greater than 2. Full details of bacterial cell growth as provided in Table 23. Figure 2 demonstrates real-time cell growth curves of the liquid bacteria cultures.

Table 23. Optical density (OD600) measurements for liquid bacteria cultures.

Sample ID	Sample name	Shake flask ID	Matrix	Condition	OD ₆₀₀ : Day 0	OD ₆₀₀ : exponential growth phase	
						Day	OD ₆₀₀
Mixed cultures							
44	SUR383/77	A	Tissue	Anaerobic	0.176	3.5	0.832
44	SUR383/77	D	Tissue	Anaerobic		3.5	1.932
215A	SUR383/77	A	Saliva	Aerobic	0.032	1	1.359
215A	SUR383/77	D	Saliva	Aerobic	0.032	1	1.294
195A	SUR400	A	Saliva	Aerobic	0.02	1	0.354
195A	SUR400	D	Saliva	Aerobic	0.013	1	0.503
197A	SUR403	A	Saliva	Aerobic	0.013	1	0.4
197A	SUR403	D	Saliva	Aerobic	0.023	1	0.122
215An	SUR383/77	A	Saliva	Anaerobic	0.192	2.5	2.042
215An	SUR383/77	D	Saliva	Anaerobic	0.188	2.5	1.522
195An	SUR400	A	Saliva	Anaerobic	0.196	2	0.416
195An	SUR400	D	Saliva	Anaerobic	0.185	2	0.415
197An	SUR403	A	Saliva	Anaerobic	0.156	2	0.363
197An	SUR403	D	Saliva	Anaerobic	0.102	2	0.335
64	SUR400	A	Tissue	Anaerobic	0.137	3	0.283
64	SUR400	D	Tissue	Anaerobic	0.139	3	0.279
39	SUR403	A	Tissue	Anaerobic	0.152	3.5	4.03
39	SUR403	D	Tissue	Anaerobic	0.15	3.5	4.79
Pure cultures							
55	402 Colony 3	A	Tissue	Anaerobic	0.179	1	2.61
55	402 Colony 3	D	Tissue	Anaerobic	0.179	1	2.07
44	77 Colony 1	A	Tissue	Anaerobic	0.15	3.5	0.229
44	77 Colony 1	D	Tissue	Anaerobic	0.171	3.5	0.268
58	78 Colony 1	A	Tissue	Anaerobic	0.185	2	0.439
58	78 Colony 1	D	Tissue	Anaerobic	0.172	2	0.57

Identity of the bacteria was confirmed by Shotgun sequencing of the mixed cultures. Sangar sequencing of the pure cultures are described in Chapter 5.1.

Shotgun metagenomic analysis of mixed bacterial cultures

Bacteria from liquid cultures from three groups of gastric cancer patients were sequenced: gastric cancer tissue and saliva (anaerobic and aerobic). Ten of the 45 cultures could not be sequenced due to low quality, however, the results available demonstrate the consistency of the bacterial profile at a genus level across the five liquid culture flasks per experiment. For example, gastric cancer bacteria from patient SUR403 grew a combination of *Veillonella*, *Streptococcus* and *Staphylococcus*; aerobic saliva cultures were composed of *Rothia*, *Bacillus*, *Neisseria*, *Haemophilus*, *Aggregatibacter* and *Streptococcus*. Two of the 5 flasks also identified *Gemella* and *Granulicatella*. A detailed profile of the mixed bacterial cultures on a species level is presented in the Table 24/Appendix 13.5.

Table 24. Mixed bacteria cultures identified by Shotgun metagenomic analysis for *in vitro* augmentation experimental set-up.

Sample type	Time interval	Taxonomy
Gastric cancer (403)	Baseline	<i>Veillonella atypica</i> , <i>Veillonella dispar</i> , <i>Staphylococcus epidermis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus vestibularis</i>
	10 minutes	<i>Veillonella atypica</i> , <i>Veillonella dispar</i> , <i>Staphylococcus epidermis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus vestibularis</i>
	30 minutes	<i>Veillonella atypica</i> , <i>Veillonella dispar</i> , <i>Staphylococcus epidermis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus vestibularis</i>
	60 minutes	<i>Veillonella atypica</i> , <i>Veillonella dispar</i> , <i>Staphylococcus epidermis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus vestibularis</i>
	120 minutes	<i>Veillonella atypica</i> , <i>Veillonella dispar</i> , <i>Staphylococcus epidermis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus vestibularis</i>
Saliva Aerobic (gastric cancer; 403)	Baseline	<i>Rothia dentocariosa</i> , <i>Bacillus amyloliquefaciens</i> , <i>Neisseria elongate</i> , <i>Haemophilus parainfluenza</i> , <i>Aggregatibacter segnis</i> , <i>Streptococcus gordonii</i> , <i>Rothia mucilaginosa</i> , <i>Neisseria flavescens</i>
	10 minutes	<i>Rothia dentocariosa</i> , <i>Bacillus amyloliquefaciens</i> , <i>Neisseria elongate</i> , <i>Haemophilus parainfluenza</i> , <i>Aggregatibacter segnis</i> , <i>Streptococcus gordonii</i> , <i>Rothia mucilaginosa</i> , <i>Neisseria flavescens</i> , <i>Neisseria sicca</i> , <i>Gemella morbillorum</i>
	30 minutes	<i>Rothia dentocariosa</i> , <i>Bacillus amyloliquefaciens</i> , <i>Neisseria elongate</i> , <i>Haemophilus parainfluenza</i> , <i>Aggregatibacter segnis</i> , <i>Streptococcus gordonii</i> , <i>Rothia mucilaginosa</i> , <i>Neisseria flavescens</i> , <i>Neisseria sicca</i> , <i>Gemella morbillorum</i> , <i>Neisseria subflava</i> , <i>Neisseria perflava</i>
	60 minutes	<i>Rothia dentocariosa</i> , <i>Bacillus amyloliquefaciens</i> , <i>Neisseria elongate</i> , <i>Haemophilus parainfluenza</i> , <i>Aggregatibacter segnis</i> , <i>Streptococcus gordonii</i> , <i>Neisseria flavescens</i>
	120 minutes	<i>Rothia dentocariosa</i> , <i>Bacillus amyloliquefaciens</i> , <i>Neisseria elongate</i> , <i>Haemophilus parainfluenza</i> , <i>Aggregatibacter segnis</i> , <i>Streptococcus gordonii</i> , <i>Rothia mucilaginosa</i> , <i>Gemella morbillorum</i> , <i>Neisseria subflava</i> , <i>Granulicatella species</i>
Saliva Anaerobic (gastric cancer; 403)	Baseline	<i>Streptococcus oralis</i> , <i>Neisseria sicca</i>
	10 minutes	<i>Streptococcus oralis</i> , <i>Streptococcus parasanguinis</i> , <i>Streptococcus mitis</i>

Sample type	Time interval	Taxonomy
Gastric cancer (400)	Baseline	<i>Bacillus paranthracis</i> , <i>Streptococcus pneumoniae</i> , <i>Streptococcus infantis</i>

	10 minutes	<i>Bacillus paranthracis</i> , <i>Streptococcus pneumoniae</i> , <i>Streptococcus infantis</i>
	30 minutes	<i>Bacillus paranthracis</i> , <i>Streptococcus infantis</i>
	60 minutes	<i>Bacillus paranthracis</i> , <i>Streptococcus pneumoniae</i>
	120 minutes	<i>Bacillus paranthracis</i> , <i>Streptococcus mitis</i>
Saliva Aerobic (gastric cancer; 400)	Baseline	<i>Neisseria sicca</i> , <i>Streptococcus oralis</i>
	10 minutes	<i>Neisseria sicca</i> , <i>Streptococcus oralis</i> , <i>Streptococcus sanguinis</i> , <i>Rothia dentocariosa</i>
	30 minutes	<i>Neisseria sicca</i> , <i>Streptococcus oralis</i>
	60 minutes	<i>Neisseria sicca</i> , <i>Streptococcus oralis</i> , <i>Streptococcus sanguinis</i>
	120 minutes	<i>Neisseria sicca</i> , <i>Streptococcus oralis</i> , <i>Streptococcus sanguinis</i>
Saliva Anaerobic (gastric cancer; 400)	Baseline	<i>Streptococcus oralis</i> , <i>Streptococcus parasanguinis</i>
	10 minutes	<i>Streptococcus oralis</i>
	30 minutes	<i>Streptococcus oralis</i>
	60 minutes	<i>Streptococcus oralis</i>
	120 minutes	<i>Streptococcus oralis</i>

Sample type	Time interval	Taxonomy
Gastric cancer (383)	Baseline	<i>Streptococcus gordonii</i> , <i>Bacillus frigoritolerans</i> , <i>Streptococcus oralis</i> , <i>Lactocaseibacillus paracasei</i> , <i>Veillonella dispar</i>
	10 minutes	<i>Streptococcus gordonii</i> , <i>Bacillus frigoritolerans</i> , <i>Streptococcus oralis</i> , <i>Lactocaseibacillus paracasei</i> , <i>Veillonella dispar</i>
	30 minutes	<i>Streptococcus gordonii</i> , <i>Bacillus frigoritolerans</i> , <i>Streptococcus oralis</i> , <i>Lactocaseibacillus paracasei</i> , <i>Veillonella dispar</i> , <i>Pauljensia odontolytica</i>
	60 minutes	<i>Streptococcus gordonii</i> , <i>Bacillus frigoritolerans</i> , <i>Streptococcus oralis</i> , <i>Lactocaseibacillus paracasei</i> , <i>Veillonella dispar</i> , <i>Pauljensia odontolytica</i>
	120 minutes	<i>Streptococcus gordonii</i> , <i>Bacillus frigoritolerans</i> , <i>Streptococcus oralis</i> , <i>Lactocaseibacillus paracasei</i> , <i>Veillonella dispar</i> , <i>Pauljensia odontolytica</i>
Saliva Aerobic (gastric cancer; 383)	Baseline	<i>Enterobacter roggenkampii</i> , <i>Enterobacter himalayensis</i>
	10 minutes	<i>Enterobacter roggenkampii</i>
Saliva Anaerobic (gastric cancer; 383)	120 minutes	<i>Veillonella atypica</i> , <i>Veillonella parvula</i> , <i>Streptococcus gordonii</i> , <i>Streptococcus oralis</i> , <i>Haemophilus parainfluenza</i> , <i>Stomatobaculum longum</i> , <i>Pauljensia species</i> , <i>Anaeroglobus micronuciformis</i> , <i>Fusobacterium periodonticum</i> , <i>Streptococcus parasanguinis</i> , <i>Porphyromonas somerae</i> , <i>Lancefieldella species</i>

Predicted metabolic pathways Bacteria profiles from gastric cancer tissue and saliva (anaerobic and aerobic cultivation) demonstrated many potential metabolic pathways with up to 100% identity to known gene clusters, associated with fatty acid VOC production. The predominant pathways identified include: pyruvate ferredoxin oxidoreductase (PFOR) II, succinate to propionate, pyruvate to acetate-formate, fumarate to succinate, propanediol degradation, formate dehydrogenase. The majority of these pathways have been previously identified in the bacteria cultivation work presented in Chapter 5.1. Predicted pathways are detailed in Table 25.

VOC augmentation Volatile organic compounds from various chemical classes were increased following glucose stimulation of the liquid culture including fatty acids, carboxylic acids, alcohols, ketones, aldehydes, alkanes and others. Short and long chain fatty acids were consistently increased across all samples; in particular 4-hydroxy butyric acid, butyric acid anhydride, pentanoic acid, 3-methyl

pentanoic acid and eicosanoic acid. Each of these compounds, excluding butyric acid anhydride, have been identified as a bacterial metabolite. All VOCs found to be augmented are detailed in Table 25.

Mixed cultures Mixed cultures established from gastric tumour tissue consistently produced eicosanoic acid following glucose stimulation of the liquid culture. In 2 of the 3 experiments, eicosanoic acid increased 1.3- and 1.4-fold at 30 minutes (OD_{600} 1.2 and 0.3 respectively). In the third case, eicosanoic acid increased 3.6-fold at 10 minutes (OD_{600} 4.4). The higher fold increases correlated with a higher optical density. Mixed saliva cultures were cultivated in both aerobic and anaerobic conditions (Figure 30).

Aerobic saliva cultures produced pentanoic and 3-methyl pentanoic acid increasing up to 10.7-fold at 120 minutes in one sample (OD_{600} 1.3), and 1.9-fold in another sample (OD_{600} 0.4). Branched butyric acid was increased in all three anaerobic saliva samples. 4-hydroxy butyric acid was increased 4.1-fold at 60 minutes and 2.8-fold at 30 minutes in two samples (Figure 31). Butyric acid anhydride increased up to 6-fold between 10 and 60 minutes between two samples.

Table 25. Summary of predicted metabolic pathways associating cancer bacteria with VOCs.

Gastric cancer bacteria	Predicted metabolic pathway	% identity to known gene cluster	VOCs (fatty acids)
Tumour (383/77): Streptococcus, Bacillus, Lactocaseibacillus, Veillonella, Pauljensenia	PFOR II pathway	100	Eicosanoic acid
	Pyruvate to acetate-formate	100	
	Fumarate to succinate	100	
	Succinate to propionate	100	
	Putrescine to spermidine	100	
	Arginine to hydrogen carbonate	100	
	Histidine to glutamate	80	
	Propanediol degradation	79	
Saliva (anaerobic; 383/77): Veillonella, Streptococcus, Haemophilus, Stomatobaculum, Pauljensenia	Pyruvate to acetate-formate	100	Pentanoic acid, Butyric acid, anhydride
	PFOR II pathway	100	
	Fumarate to succinate	100	
	Formate dehydrogenase	100	
	Acrylate to propionate	100	
	Arginine to hydrogen carbonate	100	
	Putrescine to spermidine	100	
	Glutamate to butyrate	90	
	Succinate to propionate	83	
	Propanediol degradation	75	
Saliva (aerobic; 383/77): Enterobacter	Proline to aminovalerate	66	
	PFOR II pathway	100	Pentanoic acid, 3-methyl
	Pyruvate to acetate-formate	100	
	Fumarate to succinate	100	
	Formate dehydrogenase	100	
	GlycerolP to glyceroP	100	
	Histidine to glutamate	100	
	Arginine to hydrogen carbonate	100	
	Threonine to propionate	70	
Tumour (403): Veillonella, Staphylococcus, Streptococcus	PFOR II pathway	100	Eicosanoic acid
	Pyruvate to acetate-formate	100	
	Succinate to propionate	83	
	Propanediol degradation	79	
Saliva (anaerobic; 403): Streptococcus, Neisseria	Putrescine to spermidine	100	Butyric acid, 4-hydroxy
	Arginine to hydrogen carbonate	100	
	Fumarate to succinate	100	
	Arginine to putrescine	80	
	Acetate to butyrate	50	
Saliva (aerobic; 403): Rothia, Bacillus, Neisseria, Haemophilus, Aggregatibacter, Streptococcus	Fumarate to succinate	100	
	GlycerolP to glyceroP	100	
	Formate dehydrogenase	100	
	Pyruvate to acetate-formate	100	
	Glycine cleavage	80	
Tumour (400): Bacillus, Streptococcus	Putrescine to spermidine	100	Eicosanoic acid
	Arginine to hydrogen carbonate	100	
	Pyruvate to acetate-formate	100	
	Arginine to putrescine	80	
	Histidine to glutamate	60	
Saliva (anaerobic; 400): Streptococcus	Arginine to hydrogen carbonate	100	Butyric acid, anhydride, Butyric acid, 4-hydroxy, Eicosanoic acid
	Putrescine to spermidine	80	
	Arginine to putrescine	80	
Saliva (aerobic; 400): Streptococcus	Putrescine to spermidine	100	Pentanoic acid
	Fumarate to succinate	100	
	Arginine to putrescine	80	
	Glycine cleavage	80	

	NADH dehydrogenase	76	
--	--------------------	----	--

Pure cultures Glucose stimulation of *Escherichia coli*, *Streptococcus salivarius* and *Veillonella_novel* all produced augmented levels of eicosanoic acid up to 2.8-fold. 4-hydroxybutyric acid was increased maximally in *Escherichia coli* at 5.5-fold within 10 minutes and was sustained to 3.6-fold at 120 minutes. *Streptococcus salivarius* demonstrated a maximal increase of 1.6-fold at 120 minutes (OD₆₀₀ 0.25). *Veillonella_novel* showed a 2.2-fold increase at 30 minutes (OD₆₀₀ 0.45). All fatty acids found to be increased are detailed in Table 26 and 27, Figure 32.

Table 26. Volatile organic compounds increased following glucose stimulation of bacteria cell cultures derived from tumour and saliva from oesophagogastric cancer patients.

Compound	Class of compound	Retention time 1 (min)	Retention time 2 (min)	Molecular formula	Molecular weight	Primary ions on spectra	Known bacteria metabolite
Pentanoic acid, 3-methyl-	Short chain fatty acid	18.083333	1.95018	C6H12O2	116	60, 41	Yes
Pentanoic acid	Short chain fatty acid	19.916667	1.88298	C5H10O2	102	60	Yes
Butanoic acid, anhydride	Short chain fatty acid	23.603803	0.69972	C8H14O3	158	71, 43	
Butanoic acid, 4-hydroxy-	Short chain fatty acid	16.416667	2.86326	C4H8O3	104	42, 56, 41, 86	Yes
Eicosanoic acid	Long chain fatty acid	48.711654	2.27514	C20H40O2	312	131, 117, 132	Yes
Propanoic acid, 2-methyl-, 1,2,3-propanetriyl ester	Carboxylic acid	23.583333	0.73068	C15H26O6	302	71, 43	Yes
Formic acid, dodecyl ester B	Carboxylic acid	25.673287	2.71524	C13H26O2	214	55, 43, 56	Yes
Ethanol, 2-(2-ethoxyethoxy)-	Primary alcohol	16.439568	2.35182	C6H14O3	134	45, 59, 72	-
Ethanol, 2-[2-(2-ethoxyethoxy)ethoxy]-	Alcohol	26.655323	2.33826	C8H18O4	178	45, 72, 43	-
Ethanol, 2-[4-(1,1-dimethylpropyl)phenoxy]-	Alcohol	43	2.31768	C13H20O2	208	179, 135, 107	-
Ethanol, 2-nitro-, propionate (ester)	Alcohol	14.481616	1.8648	C5H9NO4	147	74	-
2-Propanol, 1-chloro-3-methoxy-	Alcohol	25.595673	1.98888	C4H9ClO2	124	45, 75, 43	-
2-Pentanone, 5-chloro-	Ketone	6.301626	2.08404	C5H9ClO	120	43, 58, 41	Yes
1-Butanone, 3-methyl-1-phenyl-	Ketone	17.089144	2.79168	C11H14O	162	105, 77, 120, 51	Yes
Undecanal	Medium chain aldehyde	13.443517	4.4517	C11H22O	170	43, 41, 57, 55	Yes
Butane, 2,2,3,3-tetramethyl-	Alkane	3.416667	3.4479	C8H18	114	57, 56, 43	-
Hexane, 2,4-dimethyl-	Alkane	6.345749	2.03874	C8H18	114	43, 57, 41	Yes
Cyclopropane, 1,1-dimethyl-	Cycloalkane	2.526836	2.00574	C5H10	70	55, 70	-
2H-Pyran-2-one, tetrahydro-6,6-dimethyl-	Pyrones/Lactone	5	2.1342	C7H12O2	128	43, 41	Yes
m-Anisoyl amide, N-(2-phenylethyl)-N-undecyl-		40.861221	2.0676				-
3,3-Dimethyl-1,2-epoxybutane	Cyclic ether	6	2.18244	C6H12O	100	55, 70, 41	-

Table 27. Fatty acids augmented following glucose stimulation of bacteria cell cultures derived from tumour and saliva obtained from oesophagogastric cancer patients. Peak areas outputs provided by the 2D GC-MS was translated to fold change to determine augmentation.

Sample	Volatile organic compound	Time interval	Peak area	Fold change
Patient ID 77: Tumour tissue	Eicosanoic acid	Baseline	233285.9855	1.0
		10 minutes	236200.6571	1.0
		30 minutes	318501	1.4
		60 minutes	171880	0.7
		120 minutes	250592	1.1
Patient ID 77: Saliva (aerobic)	Pentanoic acid, 3-methyl	Baseline	14581	1.0
		10 minutes	18481	1.3
		30 minutes	105062	7.2
		60 minutes	137368	9.4
		120 minutes	155678	10.7
Patient ID 77: Saliva (anaerobic)	Pentanoic acid	Baseline	26754	1.0
		10 minutes	24314	0.9
		30 minutes	22467	0.8
		60 minutes	9843.0812	0.4
		120 minutes	47661	1.8
	Butyric acid, anhydride	Baseline	26333	1.0
		10 minutes	131485	5.0
		30 minutes	31256	1.2
		60 minutes	0	0
		120 minutes	50360.9949	1.9
Patient ID 400: Tumour tissue	Eicosanoic acid	Baseline	591054	1.0
		10 minutes	114487	0.2
		30 minutes	744453	1.3
		60 minutes	281646	0.5
		120 minutes	425436	0.7
Patient ID 400: Saliva (aerobic)	Pentanoic acid	Baseline	38220.9914	1.0
		10 minutes	59486.9723	1.6
		30 minutes	42145	1.1
		60 minutes	73593	1.9
		120 minutes	38473.9868	1.0
Patient ID 400: Saliva (anaerobic)	Butyric acid, anhydride	Baseline	38393	1.0
		10 minutes	182108	4.7
		30 minutes	57312.0054	1.5
		60 minutes	230384	6.0
		120 minutes	212655	5.5
	Butyric acid, 4-hydroxy	Baseline	20525	1.0
		10 minutes	75767	3.7
		30 minutes	27050	1.3
		60 minutes	84002	4.1
		120 minutes	62543	3.0
	Eicosanoic acid	Baseline	148541	1.0
		10 minutes	303022	2.0
		30 minutes	560704	3.8
		60 minutes	474932	3.2
		120 minutes	362898	2.4
Patient ID 403: Tumour tissue	Eicosanoic acid	Baseline	122719	1.0
		10 minutes	437108	3.6
		30 minutes	113049	0.9
		60 minutes	307338	2.5
		120 minutes	41676.9429	0.3
	Butyric acid, 4-hydroxy	Baseline	25208	1.0

Patient ID 403: Saliva (anaerobic)		10 minutes	17633	0.7
		30 minutes	71016	2.8
		60 minutes	24668	1.0
		120 minutes	88440	3.5
Escherichia coli (aerobic, 202)	Butyric acid, 4-hydroxy	Baseline	15576	1.0
		10 minutes	86395	5.5
		30 minutes	61109	3.9
		60 minutes	42048	2.7
		120 minutes	55941	3.6
	Eicosanoic acid	Baseline	129739	1.0
		10 minutes	161346	1.2
		30 minutes	357833	2.8
		60 minutes	259725	2.0
		120 minutes	349150	2.7
Streptococcus (58)	Butyric acid, 4-hydroxy	Baseline	72452	1.0
		10 minutes	30133	0.4
		30 minutes	34339	0.5
		60 minutes	34042	0.5
		120 minutes	114208	1.6
	Eicosanoic acid	Baseline	529610	1.0
		10 minutes	974189	1.8
		30 minutes	56032.6265	0.1
		60 minutes	48455.5131	0.1
		120 minutes	240568	0.5
	Propanoic acid, 2-methyl-, 1,2,3- propanetriyl ester	Baseline	47851	1.0
		10 minutes	22606	0.5
		30 minutes	0	0
		60 minutes	18489	0.4
		120 minutes	68906	1.4
Veillonella (58)	Eicosanoic acid	Baseline	105666	1.0
		10 minutes	35569.1279	0.3
		30 minutes	232571	2.2
		60 minutes	60094.22	0.6
		120 minutes	89917.477	0.9

Eicosanoic acid

RT₁ 48.7, RT₂ 2.27 minutes

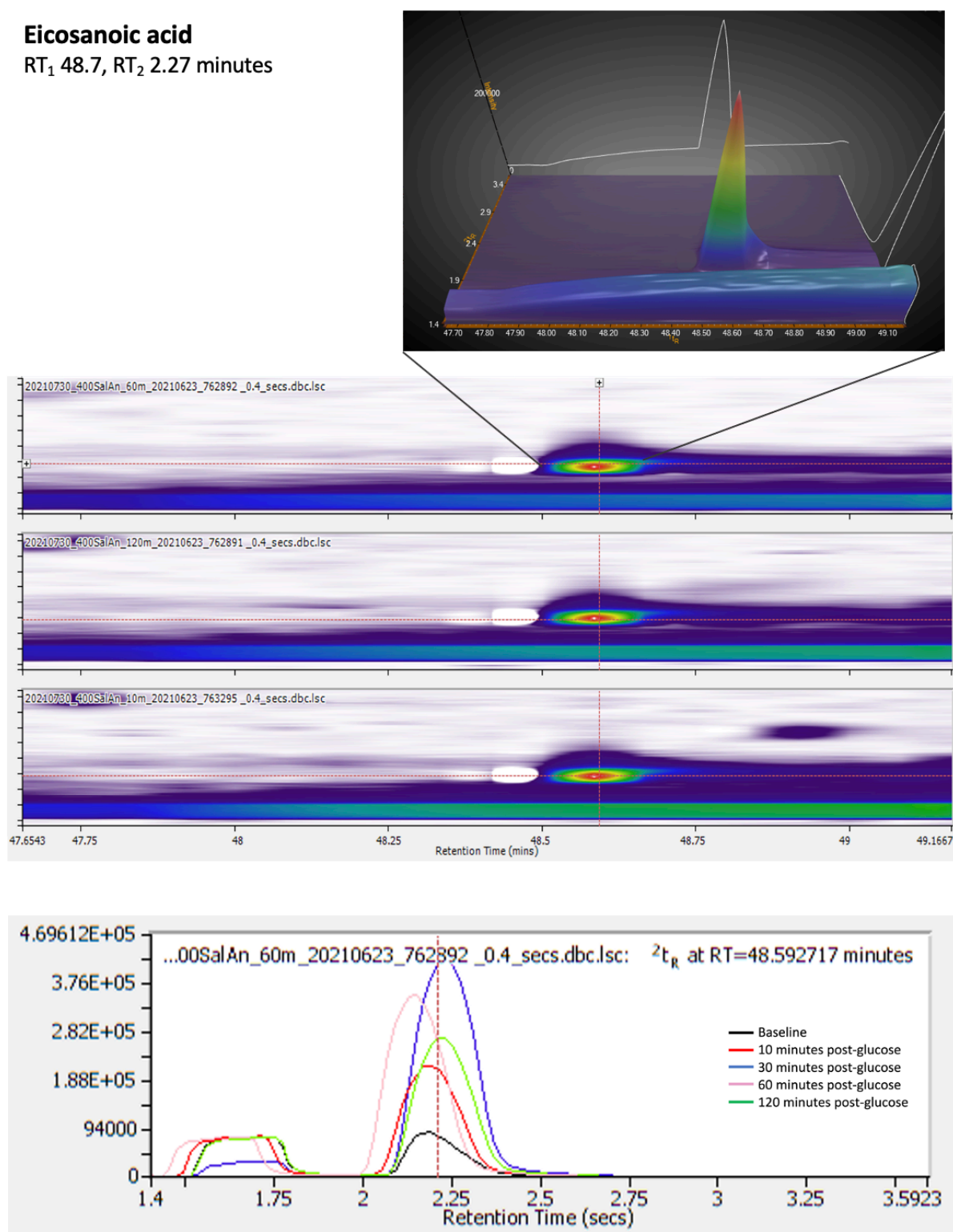


Figure 30. Chromatogram demonstrating augmentation of Eicosanoic acid following glucose stimulation of bacteria cell culture from saliva cultivated anaerobically.

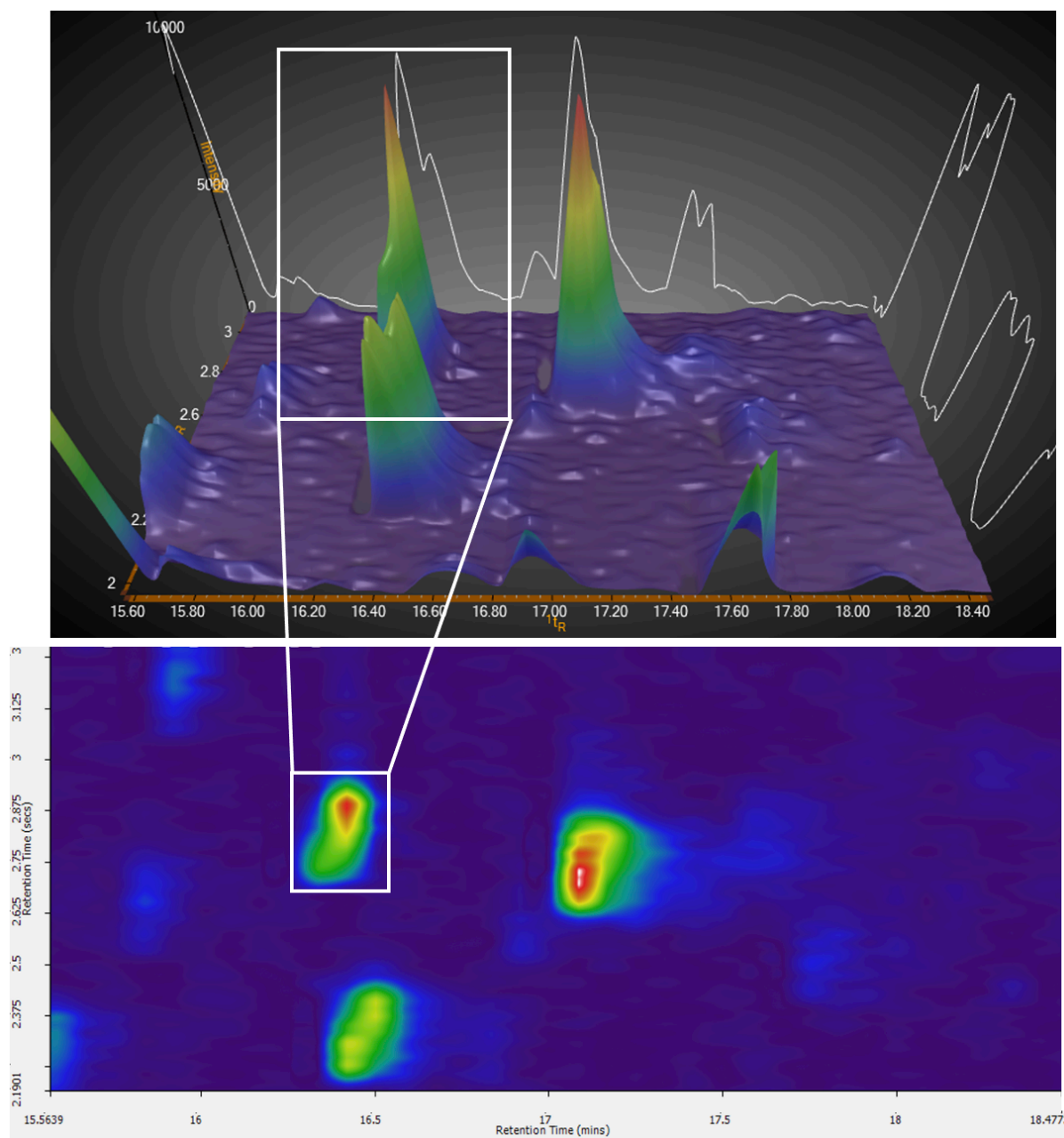


Figure 31. Chromatogram demonstrating the peak area of 4-hydroxy butyric acid detected from the headspace of *Escherichia coli* stimulated with glucose.

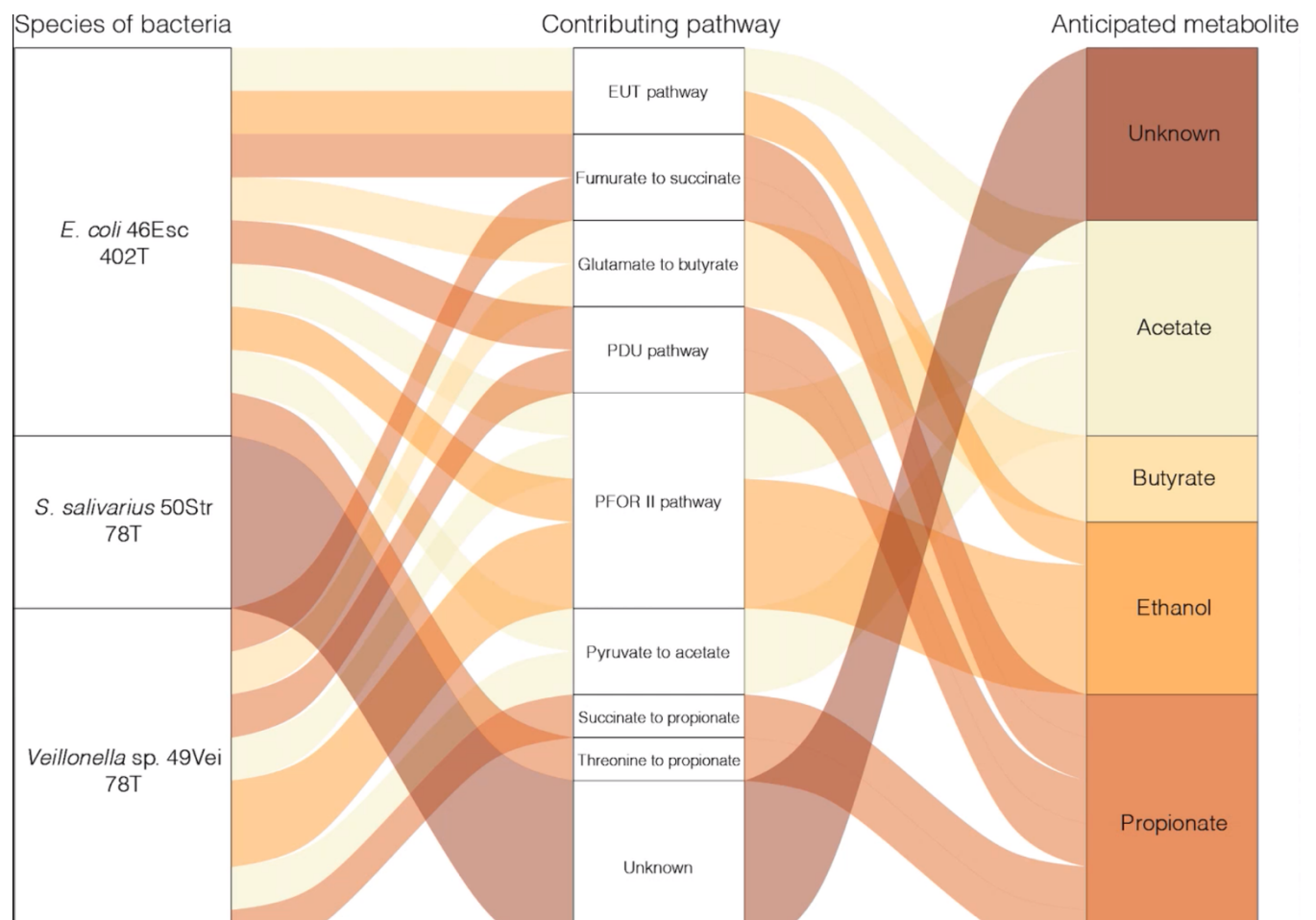


Figure 32. Sankey plot demonstrating an association between 3 pure cultures, predicted metabolic pathways and end-metabolic products; specific to oesophagogastric cancer.

9.5 DISCUSSION

The primary findings of this study are (i) augmentation of butyric acid based compounds from bacteria isolated from saliva (ii) augmentation of eicosanoic acid from bacteria cultivated from gastric cancer tissue (iii) augmentation of eicosanoic acid from three pure cultures and (iv) augmentation of short chain fatty acids (branched butyric- and pentanoic acids) from pure cultures.

Shotgun metagenomic analysis of the mixed bacterial cultures obtained from the completed augmentation experiments was performed to assess the culture setup in each flask. The bacterial profile at the species level was comparable between all the liquid culture flasks per experiment. This reflects minimal variability between culture set-up and supports the VOC augmentations detected. This taxonomic data is supported by the current literature where bacteria including *Veillonella*, *Streptococcus* and *Fusobacterium* have been associated with gastric cancer.[57, 60, 63, 65, 69, 75, 78] A similar composition of bacteria was seen across all 3 patient samples. In particular, *Veillonella* and *Streptococcus* was identified in the majority of samples. Pathway prediction describes a potential association between gastric cancer specific bacteria to target end metabolite production by glucose stimulation. The pattern of pathways identified are common to those seen in cultivation work described earlier. There is consistency across both experiments, and the suggested pathways have been shown in literature to generate fatty acids.

This study focussed and reported on the elevation of fatty acids, in keeping with the potential biomarkers relevant to oesophagogastric cancer. Both short chain and long chain fatty acids were consistently identified in the majority of bacteria samples used. Although pure butyric acid was not detected, branched butyric acids were discovered in increasingly higher levels following glucose stimulation. This is in keeping with the literature reviewed in Chapter 3 highlighting the use of glucose in bacterial metabolism to increase yield of butyric acid on an industrial scale. According to literature findings, glucose enters the glycolytic pathway, converted to pyruvate, enters the TCA cycle and downstream generates butyric acid. Butyric acid is a low molecular weight compound and volatile at temperatures above room temperature. A similar trend and pathway may be observed for increases demonstrated in pentanoic acid and its branched counterparts. A novel finding is the augmentation of eicosanoic acid which is a heavier long chain fatty acid. This was a common finding across 7 of the 11 samples.

The level of increase in fatty acids was correlated with the optical density of the liquid culture. For example, *Escherichia coli* was cultured to an OD₆₀₀ of 4.4 to generate a fold increase of 3.6, compared to *Streptococcus salivarius* which was cultured to OD₆₀₀ of 1.2 to produce a 1.3-fold increase of eicosanoic acid. The metabolic activity of each bacteria culture is unique, and therefore levels of VOC augmentation may not be consistent. However, higher optical density reflects a higher burden of bacterial cells available for metabolism. Therefore, increased metabolic activity translates to higher VOC levels.

Certain cultures such as *Escherichia coli*, or mixed aerobic populations were more easily cultivated than others. Identical culture parameters were used to emulate the human physiology. Future work may focus on optimising these parameters including prolonging culture time to increase culture yield and surface area for activity. In this study, a maximum of 3.5 days was scheduled for liquid culture growth, and all cultures were harvested during the exponential growth phase. It is important to note that in mixed cultures, different species will grow at differing rates. The diversity of the culture will influence the optimal time for exponential growth. Despite this, we observed a consistent augmentation of target VOCs following glucose stimulation.

Databases such as the Human Metabolome Database exist to correlate metabolites with literature findings. However, no such database exists for bacteria metabolites. This presents a challenge as bacteria originate from multiple sources such as humans, animals, plants and the environment, among others.

9.6 CONCLUSION

This study is crucial in demonstrating that bacteria cultivated from patients with oesophagogastric cancer, can be exploited by glucose stimulation, resulting in the augmentation of target VOCs. It has been shown to be successful in industry, and this is the first application in a clinical context. Direct manipulation of the cancer microbiome can provide a platform for novel diagnostic and treatment strategies.

10. CHAPTER 10: AUGMENTATION OF VOCS BY ANATOMICAL COMPARTMENTALISATION OF THE UPPER GASTROINTESTINAL TRACT AND THE LUNGS

10.1 INTRODUCTION

The clinical augmentation of cancer-specific VOCs has been demonstrated in mixed exhaled breath following the introduction of an oral stimulant. The consumption of glucose increases target short chain fatty acids (acetic-, butyric, and propanoic acids). However, the origin of the VOCs has not yet been ascertained. The volatile end products of metabolic degradation detected within mixed exhaled breath may be a cumulative contribution from the lungs in a systemic process, or directly from the headspace of the tumour. Anatomical compartmentalisation to isolate bronchial breath and endoluminal gas has been established by our group previously.[33] Volatile fatty acids (butyric acid $p < 0.001$, pentanoic acid $p = 0.005$) obtained directly from tumour headspace with endoluminal access were shown to be significantly elevated in oesophagogastric cancer patients compared to healthy controls. An area under the curve of 0.8 was achieved.[33] This study seeks to identify the source of VOC increase, by isolating the lung and gastrointestinal tracts. Sampling of intubated and ventilated patients is well described within the intensive care setting to investigate physiological parameters and drug therapy monitoring.[225] In particular, identification of VOCs specific to clinically relevant ventilator associated pathogens have been described.[226, 227] Non-invasive clinical tests using this principle already exist, including the urea breath test for *Helicobacter pylori* detection.[228] The glucose hydrogen breath test for small intestinal bacterial overgrowth is particularly relevant as it supports the theory that gut bacterial degradation of a stimulant can produce characteristic VOC profiles within exhaled breath via systemic absorption.[229]

10.2 AIMS

The aim of this study was to:

- (i) Assess the VOC profiles between isolated bronchial and endoluminal gaseous samples to determine the origin of VOC augmentation.
- (ii) Understand the contribution of VOCs from isolated bronchial and endoluminal samples compared to mixed exhaled breath.

10.3 METHODS

Participant selection Patients diagnosed with oesophagogastric cancer, undergoing an intraoperative oesophagogastroduodenoscopy (OGD) with either a staging laparoscopy or resectional procedure were recruited. Due to surgical protocols, all patients had been fasted overnight. Inclusion criteria included: biopsy proven oesophagogastric adenocarcinoma, aged over 18 years, able to provide informed consent, with or without neoadjuvant treatment, a clear tumour to target via the gastroscop.

Exclusion criteria included: recurrent oesophagogastric cancer, recent antibiotic use within 2 weeks, use of immunosuppressive medication, other histological sub-types including squamous cell carcinoma. Ethical approval was obtained from the London-Bloomsbury Research Ethics Committee: 18/LO/0078.

Mixed exhaled breath sampling All participants provided a single exhaled breath into a 2L polyethylene terephthalate bag (25cm x 25cm, VODCA Ltd) which transferred 500mL of breath onto two thermal desorption (TD) tubes containing 200mg Tenax and 100mg Carbograph 5 (Markes International Ltd., Llantrisant, UK). Breath was sequentially loaded onto the TD tubes with a breath sampling device (VODCA Ltd) set at a flow rate of 200mL/min for 2.5 minutes per tube. All TD tubes were conditioned at 330°C for 40 minutes (TC-20 tube conditioner, Markes International Ltd.) in a stream of pure nitrogen at 50mL/min prior to use. After breath collection, all TD tubes were immediately capped and frozen at -80°C prior to analysis.

Intraoperative sampling for isolated gas samples A system has been established to obtain isolated gas samples directly from the bronchi, and the headspace of the oesophagogastric tumour. All patients were intubated and mechanically ventilated for surgery. Total intravenous anaesthesia was used with standard ventilator settings: FiO₂ 100%, tidal volume 8mL/kg (ideal body weight), positive end expiratory pressure 5mmHg and respiratory rate 10 breaths per minute. Bronchial air was obtained via an adapter (adult patient spirometry connector, Ref 2742000, Intersurgical.com, Lithuania) placed in the ventilatory circuit between the endotracheal tube and the heat and moisture exchange filter. Caps (Codan Combi-lock, Ref 16.5285) were used to seal two ports. Less than 10cm of C-flex tubing (MasterFlex, Cole Palmer, Precision Pump tubing C-flex® Ref 06424-16) was connected to the third port and clamped with blunt forceps to maintain a closed ventilatory circuit. For sampling, the C-flex tubing was connected to the SIFT-MS instrument inlet/TD tube and the forceps released. Endoluminal sampling was performed by guiding a sampling line (V green extension, Ref 71100.20, Vygon (volume 1.0mL, L200cm, 40 bar)) through the channel of the endoscope. The luminal end of the sampling line was bevel-cut and positioned directly above the tumour. Proximally, the sampling line was connected to a drip chamber (Intrafix® SafeSet, Braun, Germany, Ref 4063000 (15µm thickness, 180cm, 2 bar)) with C-flex tubing. A 20mL leur slip syringe (BD Plastipak Syringe 300613) was used to flush the sampling line to maintain patency, after which this was connected to the instrument inlet/TD tube for sampling. The sampling set-up and processing has been detailed extensively in Nature Protocols.[230]

Two methods of capturing VOCs were used:

- Online targeted SIFT-MS. Real-time mass spectrometric analysis of target short chain fatty acids using the SIFT-MS (Voice 200, Syft Technologies). An inlet flow rate of 25mL/min was used to draw isolated gas from the bronchi/tumour headspace. Analytical timeframe of 60 seconds was observed, with a real time trace of 10 acetone peaks to reflect the standardised the ventilator settings of 10 breaths per minute. Helium carrier gas (99.996% purity) and the calibrant standard gas (48L Syft

Technologies) were used. Quality control methods have been detailed earlier in Section 4.1. Data was retrieved using the LabSyft v1.7 Software Suite.

- Offline untargeted GC-MS. Off-line analysis was performed by connecting a TD tube to the C-flex tubing of the sampling line and connecting the proximal end to a hand-pump (Pocket Pump Touch, SKC Ltd, 220-1000TC) set at a flow rate of 250mL/min for a total of 2 minutes. When 500mL of breath was loaded, TD tubes were immediately capped and frozen at -80°C until analysis. A trial sample was analysed on the polar (1D) GC-MS to confirm adequacy of the sampling parameters. The remaining samples were analysed in an untargeted fashion using the two-dimensional gas chromatography-mass spectrometer.

Quality control measures were implemented during each sample run. A judiciously selected mix of 10 standards including butyric acid, pentanoic acid, 2-hexanone, 2-octanone, decanal, hexanal, 1-nonanol, butanol, dodecane and tridecane, were made up to a 10ppm, 5ppm and 1ppm concentrations. 1µL of each stock was injected into a TD tube with nitrogen flow for 5 minutes. The quality control system employed consisted of three TD tubes (one 10ppm, one 5ppm and one 1ppm) run at the start, end, and after every 10 TD tubes containing research samples. Chromatograms were measured daily against previous quality control data to ensure integrity and no obvious shifts within the system.

10.4 RESULTS

Targeted SIFT-MS Five patients were recruited with a median age of 72 years (age range 65-79). Clinical details are provided in Table 28. Target short chain fatty acids (C2-C6) and acetone were assessed with the endotracheal and endoluminal gas. Five minutes following glucose wash, acetone increased 1.38- (2790 ppbv) and 1.3-fold (2560 ppbv) in endotracheal and endoluminal samples respectively. This increased to 1.64-fold (2650 ppbv) within the lumen at 15 minutes. Acetone is a ketone and may reflect the physiological stress exposed during an operation. Three SCFA were increased: acetic acid 1.25-fold increase (19 ppbv) at 25 minutes within the lumen, hexanoic acid 2.26-fold (3.3 ppbv) increase at 5 minutes endobronchial and pentanoic acid 1.32-fold (11.6 ppbv) increase at 5 minutes in endobronchial (Figure 33). No other changes were observed.

Table 28. Patient and clinical demographics for the intraoperative augmentation study compartmentalising two anatomical regions (endobronchial and endoluminal). GOJ = gastroesophageal junction.

Patient ID	Age	Gender	Cancer location
SUR386	75	F	GOJ – Siewert 2
SUR385	72	M	GOJ – Siewert 2
SUR400	79	F	Gastric
SUR402	65	M	GOJ – Siewert 2
SUR403	71	M	GOJ – Siewert 3

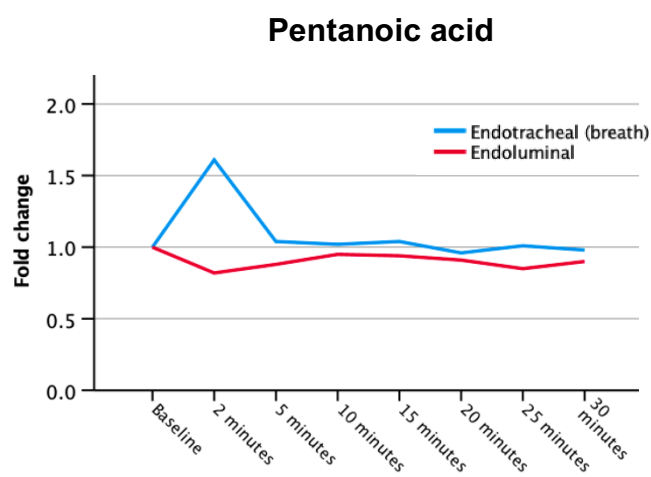
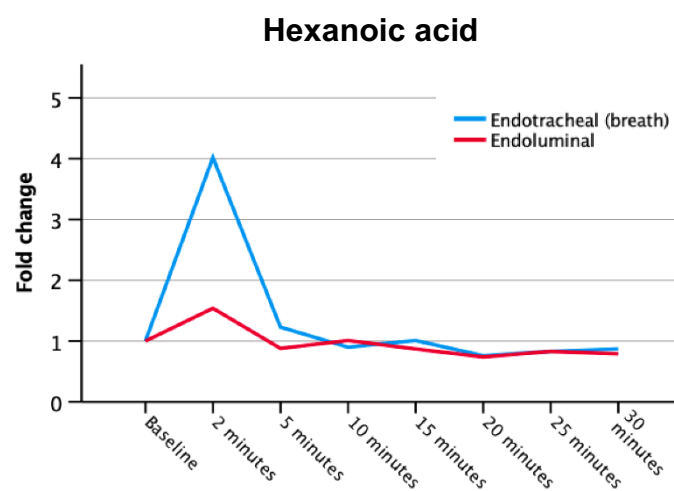


Figure 33. Fold change demonstrating the increase of hexanoic- and pentanoic acid in the endoluminal compartment following glucose wash directly over the tumour, using direct real-time analysis on the SIFT-MS.

Untargeted GC-MS One male patient aged 74 underwent a staging laparoscopy and gastroscopy for assessment of a GOJ Siewert 3 tumour (T4N1M0). The untargeted method identified an increase in compounds from the fatty acid, fatty alcohol, alcohol, aldehydes and organic acid classes. Medium chain fatty acids such as nonanoic, n-decanoic and dodecanoic acids were increased 2-3-fold in endobronchial gas, however markedly higher at 10-19 fold in endoluminal gas. For example, nonanoic acid increased maximally at 15 minutes following glucose wash. Endobronchial samples detected a 3.1-fold increase compared to endoluminal samples at 8.5-fold. Luminal gas showed increased levels as early as 5 minutes after the glucose wash (8.1-fold increase). Dodecanoic acid also demonstrated a maximal 10-fold increase at 5 minutes, reducing to 4.5-fold at 15 minutes. No changes were observed in endobronchial breath measurements. Propanoic acid, 2-hydroxy-, pentyl ester, also known as lactic acid, was also increased 20.5-fold at 15 minutes within the endoluminal tract. Alcohols and aldehydes were also elevated in luminal samples. Select compounds are detailed in Table 29 (Figures 34-36).

Table 29. Increase in volatile organic compounds in two anatomical compartments (i) endobronchial and (ii) endoluminal, following glucose wash directly over the tumour site.

Compound	Class	Retention time (minutes)	Peak intensity (Fold change)					
			Endobronchial			Endoluminal		
			Baseline	5 minutes post-glucose wash	15 minutes post-glucose wash	Baseline	5 minutes post-glucose wash	15 minutes post-glucose wash
Propanoic acid, 2-hydroxy-, pentyl ester (syn. Lactic acid)	Organic acid	5.45	4673320 (1)	259549 (0.1)	1133120 (0.2)	157524 (1)	107795 (0.7)	3221470 (20.5)
Ethanol, 2-(2-ethoxyethoxy)-	Alcohol	14.2	57416100 (1)	7515498 (0.1)	50194500 (0.9)	37885500 (1)	65091300 (1.7)	102305000 (2.7)
Dodecanal	Aliphatic aldehyde	16.1	3679260 (1)	1318920 (0.4)	5357020 (1.5)	440850 (1)	2780660 (6.5)	5062870 (11.5)
Heptanoic acid	Fatty acid	21.2	3983860 (1)	1204990 (0.3)	6011830 (1.5)	3820670 (1)	7496330 (2.0)	9715270 (2.5)
1-dodecanol	Fatty alcohol	21.4	7964640 (1)	3068089 (0.4)	11842200 (1.5)	5665420 (1)	9042350 (1.6)	15645900 (2.8)
Nonanoic acid	Fatty acid	25.3	11659900 (1)	1423830 (0.1)	36279400 (3.1)	6399200 (1)	52336100 (8.2)	54655800 (8.5)
n-decanoic acid	Fatty acid	27.2	5335700 (1)	1050174 (0.2)	11445700 (2.1)	1606233 (1)	14742500 (9.2)	29635800 (18.5)
Dodecanoic acid	Fatty acid	30.8	66959607 (1)	4747819 (0.1)	29805870 (0.4)	3131327 (1)	31350602 (10)	14429303 (4.6)

Nonanoic acid (retention time 25.3)

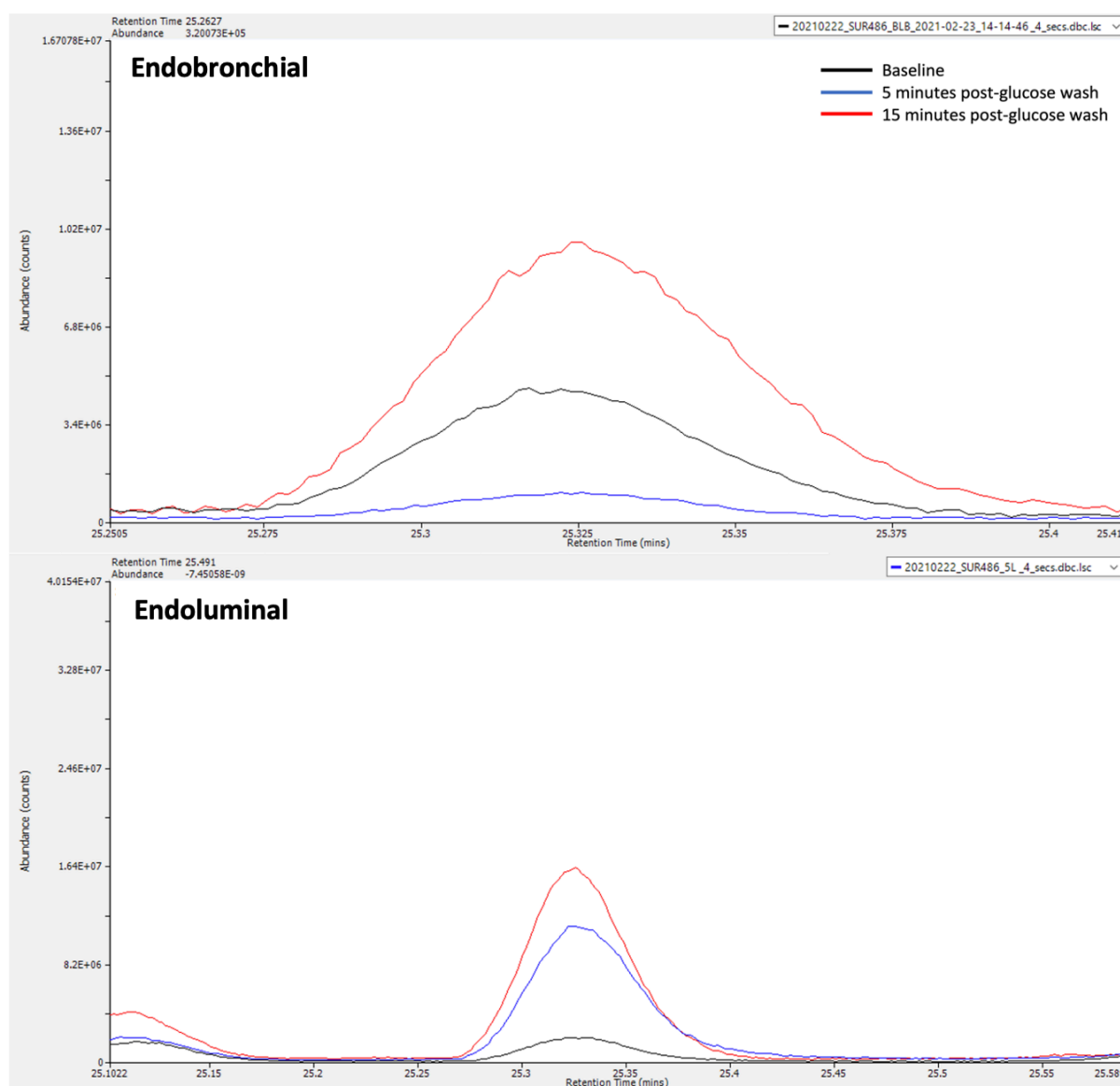
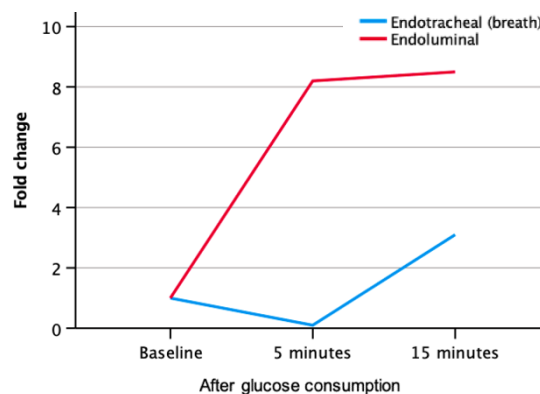


Figure 34. Chromatograms of Nonanoic acid (medium chain fatty acid) demonstrating an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).

n-decanoic acid (retention time 27.2)

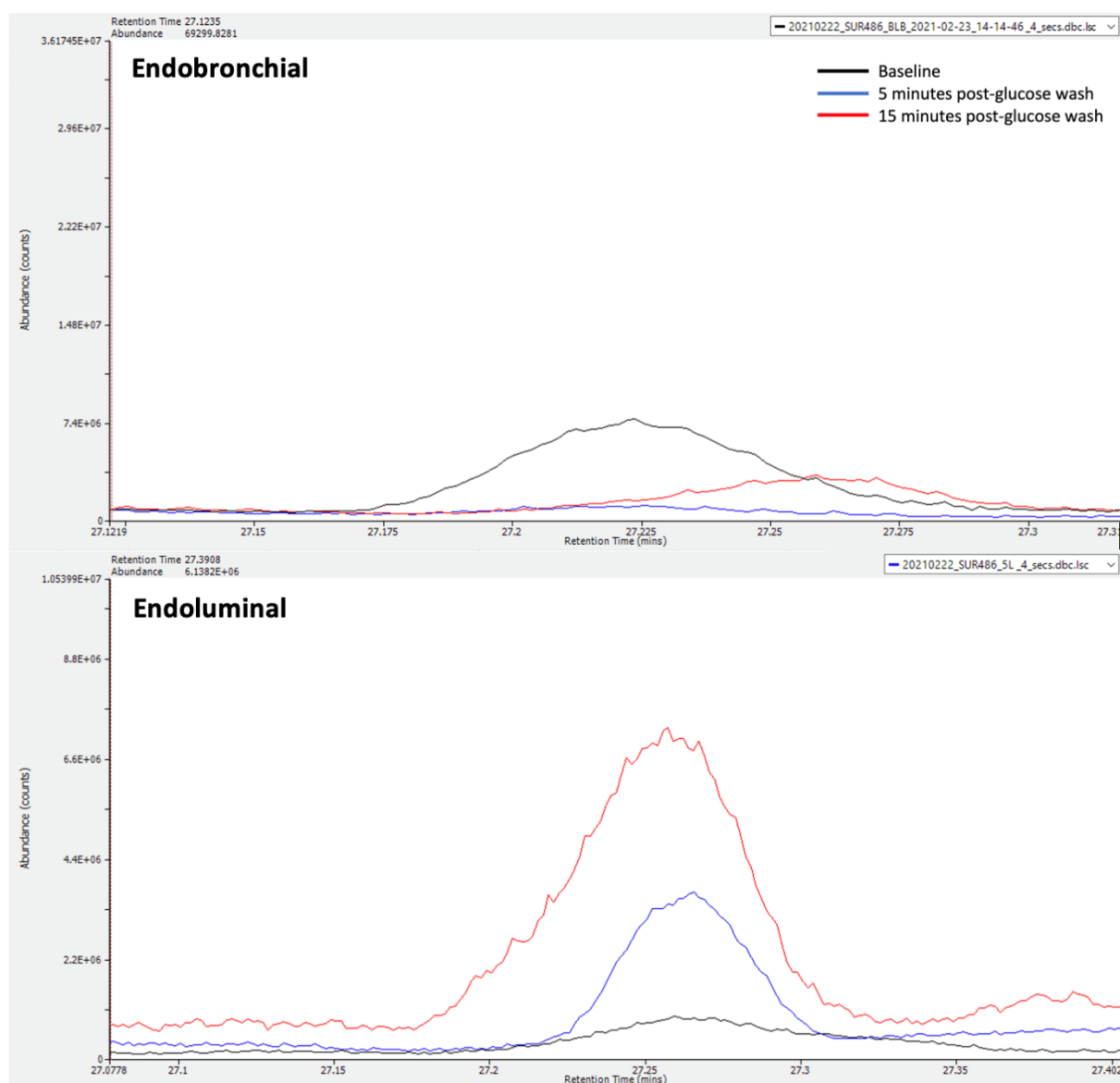
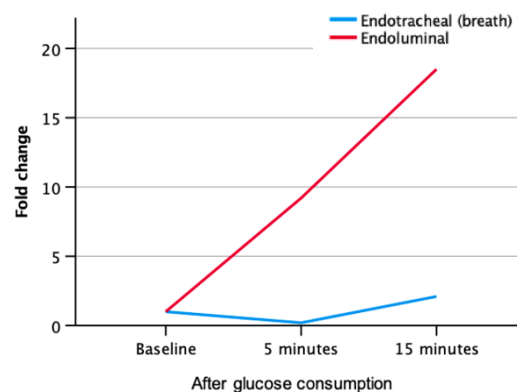


Figure 35. Chromatograms of n-decanoic acid (medium chain fatty acid) demonstrating an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time intervals (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).

Dodecanoic acid (retention time 30.8)

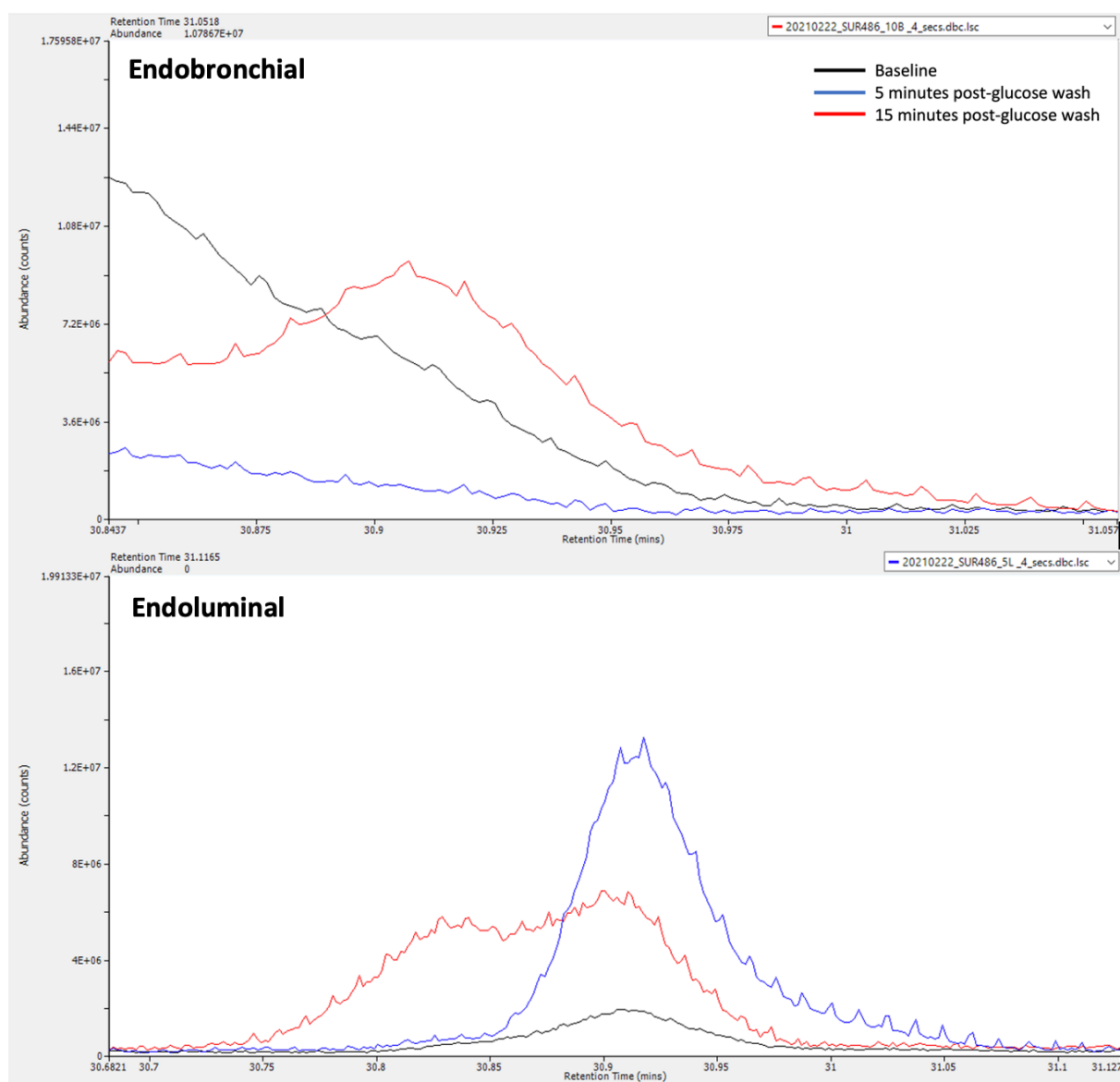
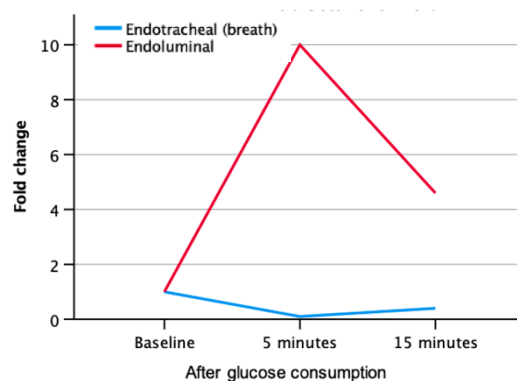


Figure 36. Chromatograms of Dodecanoic acid (medium chain fatty acid) demonstrating upregulation an increase of volatile organic compounds detected within endobronchial and endoluminal gaseous samples at three time intervals (i) baseline (black), (ii) 5 minutes after glucose wash (blue) and (iii) 15 minutes after glucose wash (red).

The protocol and parameters chosen for this study design and sample collection were considered adequate with the results generated. An increase of volatile compounds were demonstrated, suggesting the study design and volume of headspace captured was sufficient. A decision was made to use an alternative mass spectrometry platform, the 2D GC-MS for better separation and identification of the compounds. Initially, five sets of patient samples were prepared for analysis. However, samples containing luminal gas following the glucose wash produced atypical and uncharacteristic chromatograms indicating an overloaded column. The analytical parameters were changed to increase the split flow, and therefore directing less analyte volume to the column. A split ratio of 100:1 also produced similar patterns on the chromatogram. In addition, the column was found to be damaged. This identified two issues: (i) the analyte concentration was too high and (ii) volatile sugar compounds were desorbed, coated the column, and during the heating phase had polymerised and damaged the column. To preserve the sample, the run samples were recollected onto the same TD tube and transferred back to the -80°C freezer.

10.5 DISCUSSION

The principal findings of this study were: (i) *in vivo* augmentation of VOCs following glucose stimulant wash directly over the tumour (ii) primary increase of VOCs detected in the endoluminal compartment and (iii) feasibility of the study design.

These two anatomical compartments were selected to capture VOCs detected locally and systemically. VOCs emitted directly from the gastrointestinal lumen may travel proximally into the oral cavity and contribute to VOCs in breath samples. Alternatively, VOCs can be absorbed rapidly into the bloodstream and enter the systemic circulation, travel to the lungs and diffuse across the blood-alveoli interface to present in exhaled breath. Theoretically, the origin of VOCs may be isolated or from a contribution of multiple compartments.

Acetone was used as a consistent quality marker in clinical breath studies. However, the initial study using the SIFT-MS platform showed some variation in acetone concentrations. Firstly, a contribution from the ketotic stressed physiological response of the surgery, and secondly because endotracheal gas samples are not a true exhaled breath sample. The lungs are mechanically ventilated. For this reason, standardised parameters were used for ventilation based on the patients' ideal body weight.

Data obtained from the GC-MS demonstrated significant augmentation of VOCs within the endoluminal tract. Medium chain fatty acids were increased up to 12-fold.

There was one key difference identified in the results obtained from the SIFT-MS and the GC-MS. The SIFT-MS data demonstrated a primary increase in VOCs within the endobronchial compartment, compared to the GC-MS which showed an increase in VOCs within the endoluminal compartment. The

largest difference in the study design is the volume of gas between both mass spectrometry platforms. The SIFT-MS extracts 25mL of headspace per minute and therefore may not be truly reflective of the total headspace volume. This could be considered a limitation. In contrast, the GC-MS study design allowed 500mL of headspace to be collected within the TD tube. The concentration of analytes is affected by the volume of the headspace which increases following the glucose stimulant wash. It is likely the stomach capacity was higher due to gas being passed into the lumen to ensure the sampling line did not contact the stomach walls. The bronchial tree volume is on average between 300-500mL, however is dependent on patient weight. Therefore 8mL/kg tidal volume was used for ventilatory breaths. The concentration of VOCs may be different due to the different volumes; however, the purpose of the study was to detect the source of increased VOCs. The rapid nature of absorption into the blood circulation and detection in the breath is possible. For example, glucose administration sublingually in cases of hypoglycaemia, raises blood glucose levels rapidly. Therefore, the concept of glucose metabolism and absorption of VOCs systemically is plausible. This is supported by VOC detection within the lungs.

A 500mL volume was considered more representative of the headspace VOCs, opposed to 25mL. However, challenges were encountered when samples were desorbed onto the 2D GC-MS. The endoluminal samples showed evidence of analyte overload on the column, without clear separation of the compounds. There was carry-over on subsequent samples. The method was altered to increase the split ratio to reduce the sample volume and analyte concentration reaching the column and detector. A split ratio of 100:1 also produced a similar pattern with column overload. This was particularly evident for the endoluminal samples. On further investigation, the column was found to be damaged, because of likely glucose coating of the inner column, which underwent polymerisation on heating. Other causes were considered such as the suctioning of liquid onto the TD tube, however measures have been taken to avoid this by (i) ensuring the endoscopic sampling tip did not touch the luminal wall, including gastric fluid which was suctioned prior to the start of the experiment (ii) changing the sampling line and drip chamber when fluid was seen to be drawn up. Going forward, the method will be refined to obtain optimal and comparable measurements for all 15 samples. This will be facilitated by a new team member in the VOC group who has over 10 years of GC-MS experience in method development in a similar setting.

10.6 CONCLUSION

Target VOCs were increased within both anatomical compartments, however more so in the endoluminal tract than within the lung. This suggests there is a contribution of VOCs from both compartments detected within the breath. Findings will be confirmed with the remaining samples with an optimised analytical method.

11. CHAPTER 11: CLINICAL VOC AUGMENTATION WITHIN EXHALED BREATH USING NUTRITIONAL METABOLIC SUBSTRATES IN OESOPHAGOGASTRIC CANCER PATIENTS

11.1 INTRODUCTION

A novel panel of breath biomarkers have been described in oesophagogastric cancer including short chain fatty acids, aldehydes and phenols. This model has been refined and validated in a national multi-centre study demonstrating a sensitivity of 80% for a specific profile of 5-VOCs (butyric acid, pentanoic acid, hexanoic acid, butanal, decanal) in a cancer cohort of 163 patients. The challenges lie in detection of these compounds, which are highly volatile and present in part per billion per volume in exhaled breath. A proposed method of widening the gap of target compounds by augmenting the cancer-specific compounds may increase the detection rates. This is experimentally well established in multiple industries where SCFA are target end metabolites, for example in pharmaceutical, cosmetic and biofuel industries. A crossover in the end metabolite products gives an opportunity to replicate this experimentally. The successful upregulation of oesophagogastric cancer-specific bacteria following glucose stimulation has given credibility to pursue studies *in vivo*. A pilot study by our group demonstrated significantly increased levels of butyric acid in exhaled breath following consumption of an oral stimulant drink. It is important to identify links between substrates and end metabolites in a bid to define mechanistic drivers in carcinogenic pathways.

11.2 AIMS

The aim of this study was to:

- (i) Assess the dose response characteristics of exhaled breath VOC in oesophagogastric cancer patients in response to a nutrient stimulant challenge.
- (ii) Define the optimal time for maximum VOC detection.
- (iii) Formulate the optimal oral stimulant drink considering substrate selection and dose.

11.3 METHODS

Participant selection Patients diagnosed with oesophagogastric cancer and age-matched healthy controls were recruited from St Mary's Hospital through surgical and oncology clinics and endoscopy. Inclusion criteria included: patients aged 18-90 years, histologically confirmed adenocarcinoma (any stage of disease). The comparison group included patients with benign upper gastrointestinal symptoms (i.e. reflux, oesophageal dysmotility) or healthy asymptomatic controls. Exclusion criteria included: patients taking immunosuppressive medications, antibiotic use within the previous 4 weeks, co-morbidities preventing them from providing serial breath samples (i.e. severe respiratory difficulties) or unable to provide informed consent. Six patients were pragmatically selected per group (gastric adenocarcinoma, oesophageal adenocarcinoma, controls). Ethical approval was obtained from the London-Bloomsbury Research Ethics Committee: 18/LO/0078.

Exhaled breath sample collection All patients were recruited in the morning following a clear fluid diet for a minimum of six hours. After an oral rinse with water, a baseline breath sample was provided. A nutrient dissolved in 100mL of warm water was consumed, mouth rinsed with water to mitigate any potential oral contamination. Further breath samples were collected immediately after drink, and at five minute intervals for 60 minutes (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 minutes).

Selected Ion Flow Tube-Mass spectrometry All patients provided 3 consecutive exhalations over 60 seconds via a disposable mouthpiece directly into the SIFT-MS, described in Section 4.1. Target VOCs from four groups (short chain fatty acids, aldehydes, phenol-alkanes and hydrocarbons) were selected.

Nutrient stimulant substrates Substrates were judiciously selected based on the target VOCs and the potential intrinsic metabolic pathways involved. Six experiments were designed with single and combination substrates with doses determined according to the National Academics Dietary Reference Intake (Table 30). All substrates were stored in their powdered/liquid form at room temperature 15-25°C (Sigma Aldrich) until required.

Table 30. Nutrient stimulants selected to increase target VOCs.

Substrate	Biochemistry	Molecular structure	Pathway	End metabolite
Carbohydrates				
Experiment 1: Glucose 25g	Glucose: Monosaccharide, hexose	C ₆ H ₁₂ O ₆	Glycolytic	Short chain fatty acids Aldehydes
Experiment 2: Glucose 25g + Lactose 25g	Lactose: Disaccharide	C ₁₂ H ₂₂ O ₁₁		
Amino acids				
Experiment 3: Tyrosine 2g	Tyrosine: Non-essential amino acid	C ₉ H ₁₁ NO ₃	Shikimate	Phenol P-cresol
Experiment 4: Tyrosine 2g + Phenylalanine 3g + Glutamic acid 2.1g	Phenylalanine: Essential amino acid (precursor to tyrosine)	C ₉ H ₁₁ NO ₂		
	Glutamic acid: Non-essential amino acid	C ₅ H ₉ O ₄ N		
Lipids				
Experiment 5: Glycerol 25g	Glycerol: Polyol compound	C ₃ H ₈ O ₃	Glycolytic	Short chain fatty acids Aldehydes
Carboxylic acids				
Experiment 6: Citric acid 1.4g + Glucose 25g	Citric acid: Intermediate in the citric acid cycle	C ₆ H ₈ O ₇	Citric acid cycle	Short chain fatty acids Aldehydes

11.4 RESULTS

A total of 77 patients were recruited with a median age of 68 (IQR: 61-73) and 45 male participants (58%). There were no significant differences in baseline characteristics across the three patient groups: gastric adenocarcinoma (n=21), oesophageal/gastro-oesophageal junction adenocarcinoma (n=18) and healthy controls (n=38) (Table 31).

Acetone was used as a quality control marker to assess for an adequate breath sample. It is considered a systemic marker of ketosis, detected in the breath after a period of fasting. Acetone was present at levels over 200 ppbv consistently for all breath provided.

Table 311. Baseline characteristics between the three patient groups.

	Gastric adenocarcinoma	Oesophageal adenocarcinoma	Controls
Cases, n (%)	21 (27)	18 (23)	38 (49)
Age, y (IQR)	69 (62-74)	70 (60-73)	68 (61-72)
Gender (M:F)	15:6	17:1	13:25
Clinical staging, n (%)			
T1	1 (4)	1 (6)	-
T2	1 (4)	1 (6)	-
T3	3 (14)	13 (72)	-
T4	10 (48)	3 (17)	-
N0	4 (19)	11 (61)	-
N1	5 (24)	4 (22)	-
N2	6 (29)	3 (17)	-
N3	2 (10)	0	-
M0	12 (57)	3 (17)	-
M1	8 (38)	15 (83)	-
Tumour Differentiation			
Poor	8	4	-
Moderate	3	7	-
Well	0	0	-
Treatment Pathway			
Curative	14 (67)	13 (72)	-
Palliative	7 (33)	5 (28)	-
Chemo/radiotherapy			
Yes	15 (71)	11 (61)	-
No	6 (29)	7 (39)	-
Controls: diagnosis			
Gastro-oesophageal reflux	-	-	13 (34)
Gastritis	-	-	1 (3)
Hernia	-	-	2 (5)
Hiatus hernia	-	-	1 (3)
Oesophageal dysmotility	-	-	2 (5)
Benign dysphagia	-	-	1 (3)
Gallstones	-	-	1 (3)
Other	-	-	17 (45)

Single substrates

Glucose Glucose consumption produced marked increases in target VOCs. Within 5 minutes, butyric acid was elevated at 10.5-fold (median 5- to 34 ppbv) and 5-fold (median 4.4- to 12.3 ppbv) in gastric adenocarcinoma and oesophageal adenocarcinoma, respectively. Healthy controls showed a 2.7-fold increase from a baseline of 5.6 ppbv to 9.6 ppbv. Levels returned to baseline within 20 minutes. In addition, propanoic acid and acetic acid were increased in gastric cancer with a fold increase of 3.5 (8.8- to 38 ppbv) and 2.5 (16- to 39 ppbv) respectively. Select aldehydes; butanal, pentanal, and propanal were augmented between 5-15 minutes in gastric cancer, albeit at lower concentrations. The most marked increases were seen with pentanal (1.6-fold) and propanal (1.6-fold). No changes were observed with the oesophageal adenocarcinoma cohort. A new finding of increased octanal in the control group only (1.4-fold increase) was observed (Table 32, Figure 37).

Glycerol Glycerol, a viscous polyol compound, demonstrated similar augmented trends to glucose. Both oesophageal and gastric cancer groups showed an increase of butyric acid to 2-fold (10- to 19 ppbv) and 1.4-fold (5- to 8.7 ppbv) respectively. Concentrations of propanoic acid were also higher in oesophageal cancer (1.8-fold increase; 46- to 108 ppbv) compared to gastric cancer (1.5-fold; 10- to 19 ppbv). Acetic acid showed a 1.5-fold increase in oesophageal cancer. Aldehydes such as hexanal (1.7-fold) and propanal (1.7-fold) were raised in oesophageal cancer only. Pentanal concentrations were increased in oesophageal (1.3-fold) and gastric (1.5-fold) cancers. The optimal time point was seen at between 45-60 minutes (Table 33, Figure 38).

Tyrosine No apparent changes were observed with the consumption of tyrosine (Appendix 13.6).

Table 322. Glucose consumption: Comparison in the median concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.

	Control (n=6)			Oesophageal Cancer (n=6)			Gastric Cancer (n=6)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-glucose: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glucose: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glucose: Concentration (ppbv)	Fold change	
Short chain fatty acids										
Acetic acid	38.03	31.37	0.66	19.58	23.63	1.11	16.63	39.0	2.49	5-10
Butyric acid	5.57	9.64	2.72	4.40	12.34	1.61	4.96	34.18	10.53	5-10
Hexanoic acid	0.64	0.52	0.87	1.00	0.65	0.63	0.76	0.63	0.87	5-10
Pentanoic acid	2.45	2.38	0.94	3.62	1.88	0.67	2.86	2.69	1.16	25-30
Propanoic acid	15.89	17.14	1.17	17.15	30.20	1.33	8.84	38.22	3.52	5-10
Aldehydes										
Acetaldehyde	16.02	15.25	1.04	22.91	20.48	1.02	15.12	14.68	0.88	5-10
Butanal	2.46	2.21	1.03	1.99	2.33	1.22	2.56	2.93	1.29	5-10
Decanal	0.66	0.59	0.97	0.60	0.81	1.24	0.65	0.67	1.06	5-10
Heptanal	0.60	0.53	0.86	0.80	0.76	0.93	0.98	0.84	0.88	5-10
Hexanal	2.71	3.09	1.06	2.83	2.77	0.95	2.90	3.42	1.06	5-10
Nonanal	1.07	1.56	1.23	1.49	1.56	1.01	1.83	1.64	0.93	5-10
Octanal	0.43	0.58	1.40	0.79	0.67	0.83	1.02	0.83	0.90	5-10
Pentanal	1.17	1.05	1.00	1.17	1.08	0.78	1.05	1.84	1.59	15
Propanal	9.67	9.29	1.06	9.26	11.47	1.19	5.18	14.37	1.64	5-10
Phenol-alkanes										
1-hydroxy-4-ethylbenzene	0.36	0.38	0.86	0.48	0.47	1.11	0.36	0.51	1.21	5-10
Decane	2.45	2.48	1.26	2.90	2.90	1.03	2.75	2.84	0.90	5-10
Dodecane	0.66	0.51	0.84	0.65	0.59	1.25	0.55	0.54	1.10	5-10
P-cresol	0.97	0.84	1.03	1.16	1.24	1.11	1.09	0.93	1.25	5-10
Phenol	7.12	4.53	0.59	4.74	3.19	0.70	2.17	1.56	0.92	5-10

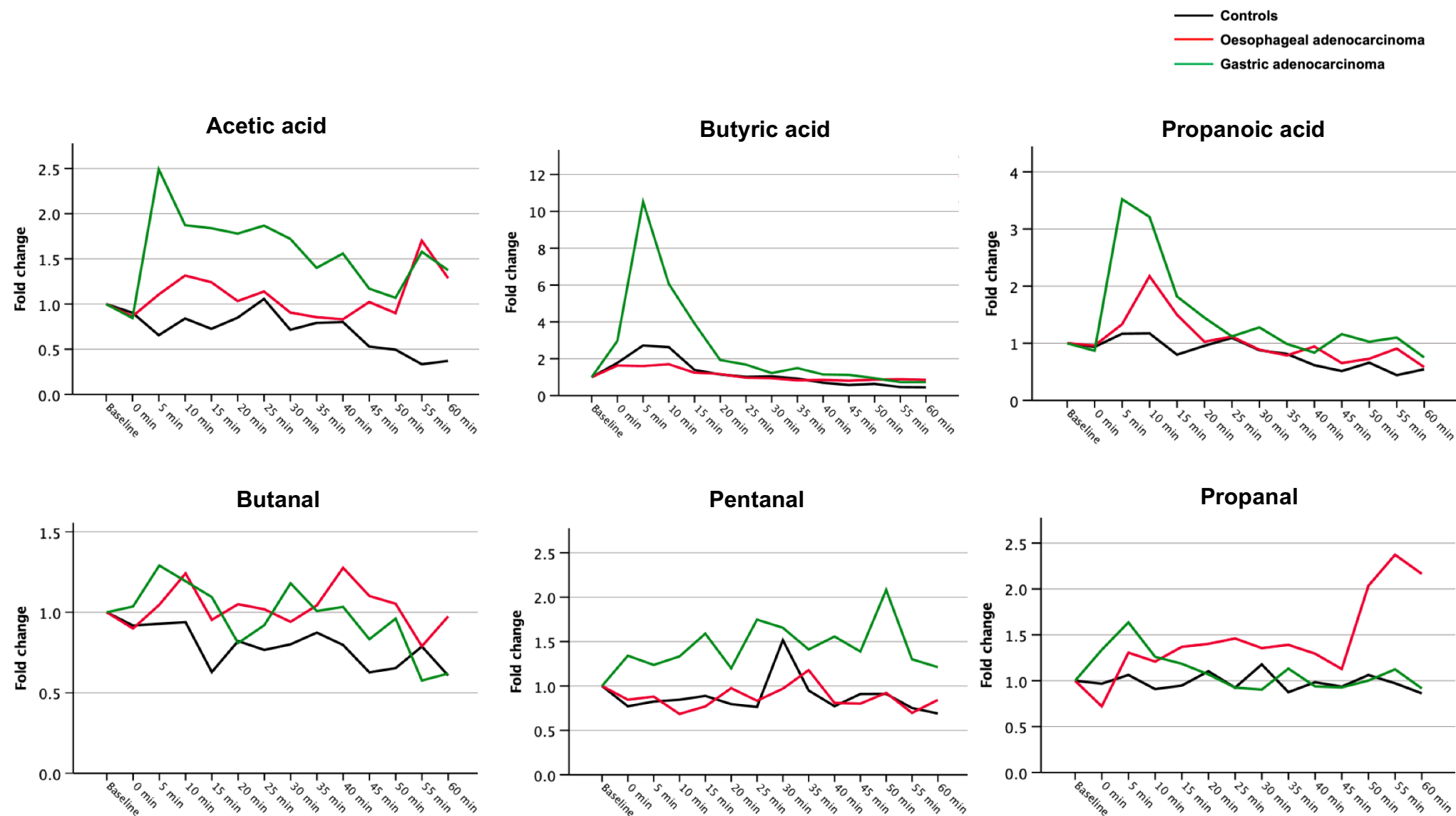


Figure 37. Glucose consumption: Fold increase of target short chain fatty acids (acetic-, butyric- and propanoic acids) and aldehydes (butanal, pentanal and propanal) between three cohorts; oesophageal adenocarcinoma, gastric adenocarcinoma and healthy controls.

Table 333. Glycerol consumption: Comparison in the median concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.

	Control (n=6)			Oesophageal Cancer (n=6)			Gastric Cancer (n=6)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-glycerol: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glycerol: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glycerol: Concentration (ppbv)	Fold change	
Short chain fatty acids										
Acetic acid	35.68	26.90	0.62	54.77	101.35	1.50	20.99	33.60	1.09	60
Butyric acid	7.46	8.59	1.12	10.55	18.93	2.02	4.99	8.67	1.43	60
Hexanoic acid	1.56	1.13	0.52	2.04	1.43	0.49	1.95	0.98	0.70	60
Pentanoic acid	3.53	3.75	0.83	5.49	7.87	1.05	2.86	2.49	0.86	60
Propanoic acid	24.70	26.95	0.72	45.95	107.93	1.75	10.19	19.07	1.46	60
Aldehydes										
Acetaldehyde	25.47	32.05	0.88	51.00	19.30	0.80	23.72	17.85	0.83	60
Butanal	3.39	4.44	0.87	9.02	6.49	1.00	2.78	2.21	1.05	60
Decanal	1.31	1.18	0.97	2.53	1.46	0.77	1.17	0.95	1.05	60
Heptanal	2.28	1.33	0.58	2.48	1.40	0.76	1.60	1.20	0.65	60
Hexanal	6.28	6.93	1.17	7.77	11.56	1.72	4.42	5.52	1.09	40
Nonanal	3.07	2.97	0.98	3.62	3.71	1.09	3.36	2.93	1.09	60
Octanal	1.49	1.35	0.95	1.91	1.68	1.58	1.25	1.39	0.87	40
Pentanal	2.62	1.39	0.63	2.37	3.71	1.27	1.30	1.63	1.46	55
Propanal	15.77	12.51	0.77	16.93	17.83	1.71	7.38	5.67	0.98	55
Phenol-alkanes										
1-hydroxy-4-ethylbenzene	0.72	1.09	1.47	1.14	0.88	0.99	0.56	0.56	1.47	60
Decane	5.33	4.61	0.92	6.54	7.33	1.13	5.83	4.89	0.81	60
Dodecane	1.03	0.75	0.86	1.02	1.30	1.09	0.84	1.05	0.95	60
P-cresol	1.71	1.46	0.71	2.23	2.41	1.04	0.96	0.83	0.98	60
Phenol	4.21	2.64	1.43	4.73	7.71	1.36	4.18	3.68	0.98	60

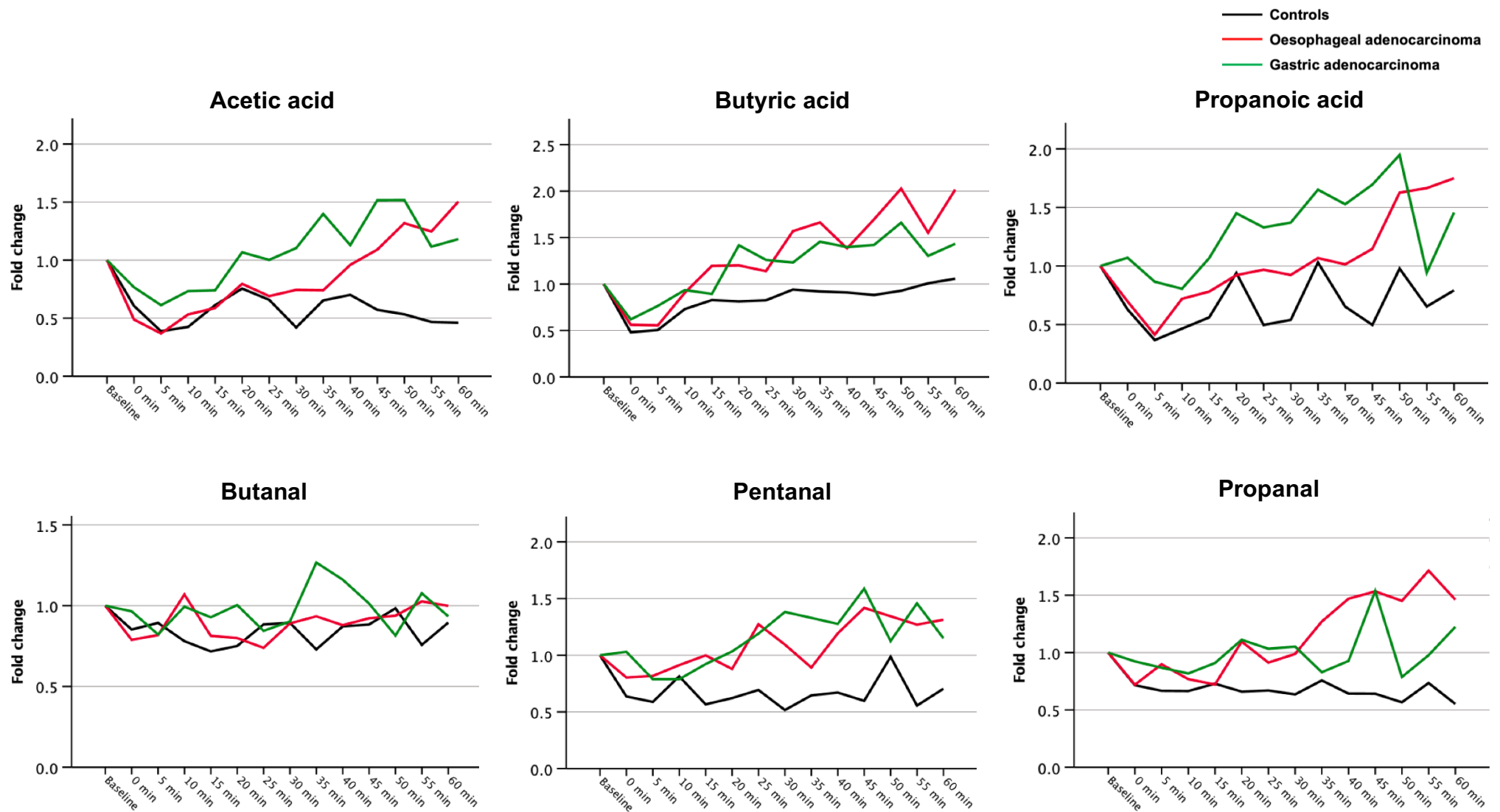


Figure 38. Glycerol consumption: Fold increase of target short chain fatty acids (acetic-, butyric- and propanoic acids) and aldehydes (butanal, pentanal and propanal) between three cohorts; oesophageal adenocarcinoma, gastric adenocarcinoma and healthy controls.

Combined substrates

Carbohydrates (glucose + lactose) An increase in the SCFA of interest (acetic-, butyric- and propanoic acids) were observe. In comparison to glucose alone, the optimal time point for augmentation was seen between 10-30 minutes. Butyric acid was increased 3.8-fold in controls, 4.5-fold in oesophageal cancer and 5.6-fold increase in gastric cancer. A similar pattern was seen with acetic acid concentrations; 1.7-fold controls, 2.4-fold oesophageal cancer, 4.3-fold gastric cancer. The scale of augmentation was reduced compared to glucose alone. Butanal showed a modest increase in the cancer groups (1.4-1.6-fold) in comparison to no change in the healthy controls (Appendix 13.7).

Amino acids (tyrosine, phenylalanine, glutamic acid) A single compound, p-cresol (4-methyl-phenol) was found to be increased in oesophagogastric cancer at 1.5-fold (0.9- to 1.4 ppbv) after 20 minutes of consumption. No changes were seen within the control group. (Table 34, Figure 39).

Glucose + citric acid No sustained differences in VOC increase was noted across all VOCs of interest. Butyric acid was increased to the similar levels between controls (5.4-fold; 2.9- to 15 ppbv) and oesophagogastric cancer (6.2-fold; 7.6- to 28 ppbv) (Appendix 13.8).

Glucose oral rinse With increasing evidence that the oral microbiome is distinctive in cancer, coupled with the outcome of SCFA and aldehyde augmentation, an oral rinse with glucose water was trialled. Butyric-, acetic- and propanoic acids were significantly increased with no clear differences between the cancer and control groups. In contrast, butyric acid was 8-fold higher in this experiment (11-fold) compared to glucose water drink (3-fold) in healthy controls. This is likely to be attributed to the higher concentration of glucose within the oral cavity, opposed to a superficial coating on consuming a drink (Appendix 13.9).

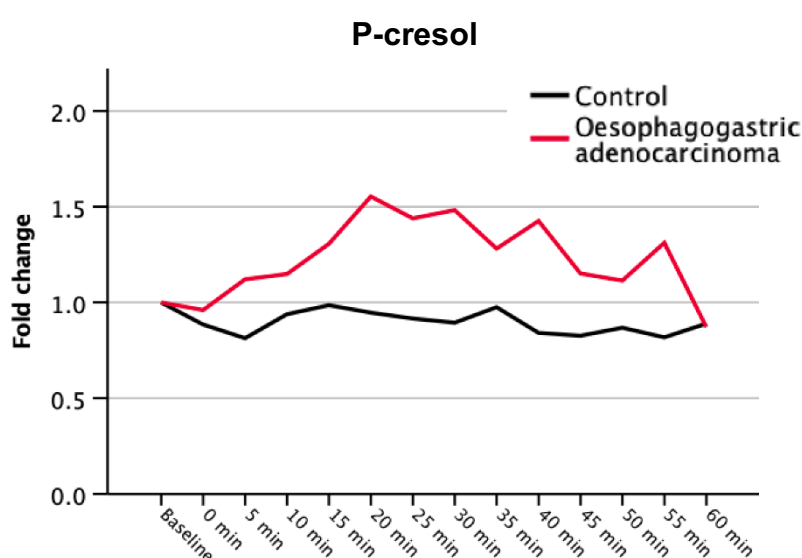


Figure 39. Combined amino acids consumption: Fold increase of p-cresol between two cohorts; oesophagogastric adenocarcinoma and healthy controls.

Table 344. Combined amino acids consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls and oesophagogastric adenocarcinoma.

	Control (n=6)			Oesophagogastric Cancer (n=7)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-amino acids: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-amino acids: Concentration (ppbv)	Fold change	
Short chain fatty acids							
Acetic acid	25.7	24.7	0.67	24.6	29.2	0.99	30
Butyric acid	5.10	3.74	0.88	8.13	7.76	1.15	30
Hexanoic acid	0.71	0.65	0.94	1.10	0.89	0.81	30
Pentanoic acid	2.08	1.74	1.00	2.24	1.53	0.84	30
Propanoic acid	13.8	9.95	1.10	11.7	11.5	0.91	30
Aldehydes							
Acetaldehyde	17.0	15.1	1.00	30.1	16.7	0.59	45
Butanal	2.39	2.29	1.28	2.48	2.31	0.88	35
Decanal	0.64	0.60	1.05	0.69	0.83	1.41	30
Heptanal	1.23	0.65	0.82	1.20	0.89	0.84	35
Hexanal	2.73	3.47	1.65	2.52	3.60	1.32	55
Nonanal	1.68	2.27	1.31	2.21	2.32	1.11	50
Octanal	0.82	0.54	0.94	0.88	0.74	0.67	35
Pentanal	1.06	0.69	1.42	0.98	1.01	1.00	60
Propanal	12.2	9.16	0.86	8.55	10.3	1.18	50
Phenol-alkanes							
1-hydroxy-4-ethylbenzene	0.64	0.56	1.24	0.65	0.62	1.28	20-25
Decane	3.59	4.92	1.19	3.87	5.62	1.46	45
Dodecane	0.52	0.64	1.39	0.76	0.84	1.30	45
P-cresol	1.26	0.98	0.95	0.93	1.40	1.51	20
Phenol	3.15	5.75	2.13	1.59	2.06	1.83	20-30

11.5 DISCUSSION

The purpose of this study was to (i) identify nutritional substrates capable of augmenting target VOCs and (ii) define the optimal time points for augmentation. A previous pilot study by the group demonstrated an increase of short chain fatty acids, in particular butyric acid, at 30 minutes after consumption of an oral stimulant drink. This study further refined the parameters for optimal augmentation of target VOCs in oesophagogastric cancer. The primary findings of this study are augmentation of: (i) butyric-, propanoic- and acetic acid 5 minutes after glucose consumption (ii) butanal and propanal after 60 minutes of glucose consumption (iii) butyric acid, butanal and propanal after 30 minutes of glycerol consumption and (iv) p-cresol 20 minutes after consumption of combined amino acids. Other important findings were increasing the sugar dosage and oral rinsing with glucose eliminated the discrimination in VOC detection between cancer and control cohorts.

Initially breath was analysed every 5 minutes for 2 hours, however, early on it was clear that the response to glucose had returned to baseline values within an hour. All subsequent analyses were performed for one hour. Glycerol solution demonstrated a gradual increase up to 60 minutes. Glycerol is a more complex substrate compared to glucose and amino acids. It requires degradation before it can enter a metabolic cycle, where glucose in its simple form can integrate immediately.

Identifying the optimal timing for augmentation was crucial to design a future clinical study. Glucose was degraded rapidly, and short chain fatty acids were increased after 5 minutes of consumption. In particular, acetic-, butyric- and propanoic acids were increased 3-, 10.5, and 4-fold respectively. These compounds have been strongly associated with the oral cavity in the context of periodontal disease. Dental plaques and bacteria within periodontal pockets produce butyric acid. An oral water rinse after the metabolic drink partially eliminates VOC production from the oral cavity. However, given earlier methodology studies, it is confirmed that butyric acid is produced within the mouth to some degree. This is important because the oral microbiome has been implicated in potential cancer diagnostics, with a unique microbial profile. The concept of VOC augmentation by microbial metabolism of substrates may be linked with the oral microbiome. The cancer microbiome is reported to be altered along the upper gastrointestinal tract, including the oral cavity, tumour, and adjacent to tumour site. Consuming an oral metabolic drink coats the entire upper gastrointestinal tract presenting an opportunity for metabolic reactions from the mouth to the stomach. Exposing the oral cavity to glucose alone demonstrated a loss of discrimination between cancer and control cohorts. This was likely due to higher concentration within a small cavity. This suggests that the effects of the oesophagus and stomach are important in the augmentation process.

Glycerol is a polyol complex compound which requires degradation before it enters the metabolic cycle. This may suggest why the optimal time point for VOC increase is a gradual increase up to 60 minutes. The oesophageal cancer group demonstrated a greater response with glycerol which may be attributed to its viscous composition and thus prolonged contact of the oesophageal tract and tumour.

This clinical study supports the findings of the bacterial augmentation study *in vitro* which demonstrated that bacteria can be exploited to increase target VOCs with the introduction of a glucose stimulus. Introduction of a simple nutrient stimulates bacteria producing a higher VOC concentration. Although this is novel in the medical field, bio-fermentation is well-established in multiple industries, predominantly the biofuel industry. The biofuel industry optimises the reaction parameters to increase the yield of butyric acid from glucose. The fundamental principle of this work is similar. However, this raises the question of the metabolic processes linking the introductory stimulus to the end-metabolic products. Literature potentially offers an explanation for this link. The glycolytic pathway may be responsible for the production of short chain fatty acids, in particular acetic-, butyric- and propanoic acids. Glucose enters the glycolytic pathway directly, degrades to pyruvate and enters the citric acid cycle for end metabolite production (Figure 40). Glycerol is transported into the cell via the GlpF, a glycerol uptake facilitator protein, by means of diffusion. At this point, glycerol undergoes a dehydrogenase process anaerobically to dihydroxyacetone, or a phosphorylation process aerobically to glycerol-3-phosphate before final conversion to dihydroxyacetone phosphate (DHAP). DHAP directly enters the glycolytic pathway (Figure 41). The shikimate pathway has been described in bacteria for the conversion of tyrosine to phenol and p-cresol. Two pathways have been reported for this conversion (i) the tyrosine phenol lyase enzyme is responsible for the direct conversion and (ii) tyrosine aminotransferase B (TyrB) converts tyrosine to 4-hydroxyphenylpyruvate followed by multiple enzymatic reactions to 4-hydroxybenzoate and finally phenol. Similarly, tyrosine is converted to cresol (i) directly by tyrosine lyase (ThiH) and (ii) indirectly via TyrB and multiple enzymatic reactions to 4-hydroxyphenylacetate to cresol (Figure 42). Further investigation needs to elucidate the pathways.

11.6 CONCLUSION

Oesophagogastric cancer remains poorly diagnosed with low detection rates. An augmentation of candidate biomarkers specific to oesophagogastric cancer presents a novel test for early detection. This is a niche area which can be exploited in a commercial setting. However, it primarily offers a safe test with negligible adverse effects. This study has defined important parameters to design a future trial to confirm these findings.

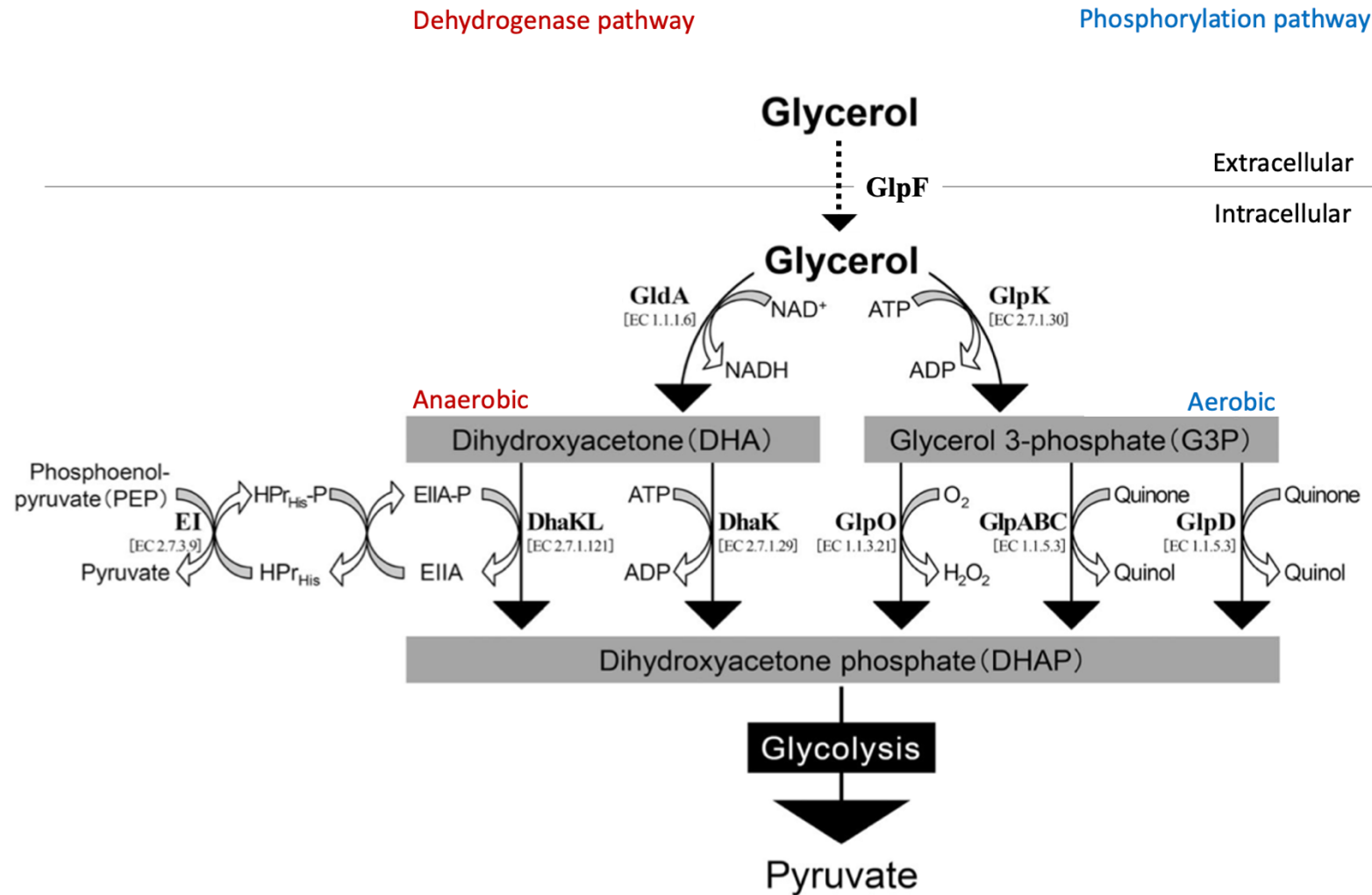


Figure 40. Potential metabolic pathway for glycerol metabolism; (i) Dehydrogenase pathway (ii) Phosphorylation pathway. Adapted from [231-233]

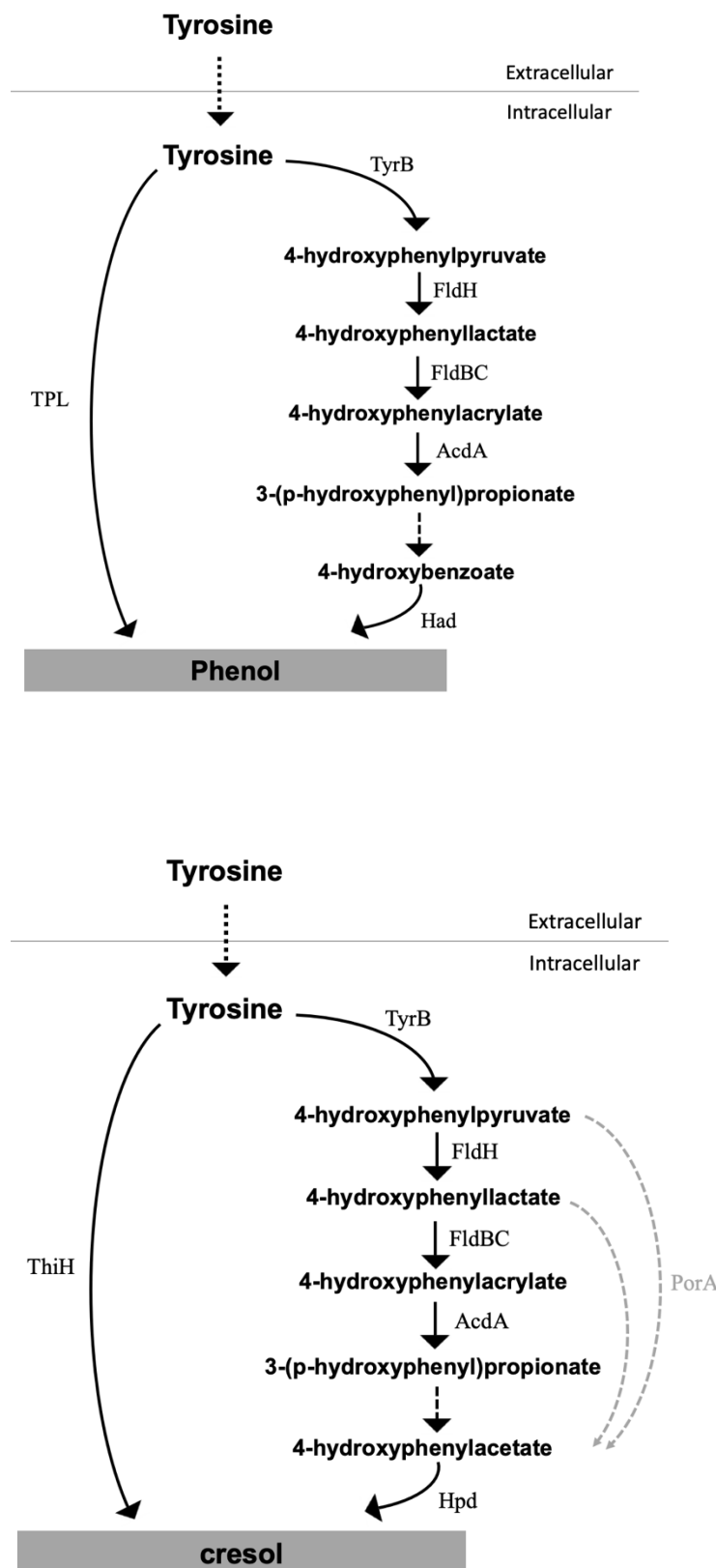


Figure 41. Potential metabolic pathway for tyrosine metabolism to (i) phenol and (ii) cresol. TPL = tyrosine phenol lyase, ThiH = tyrosine lyase, TyrB = tyrosine aminotransferase. Adapted from [234]

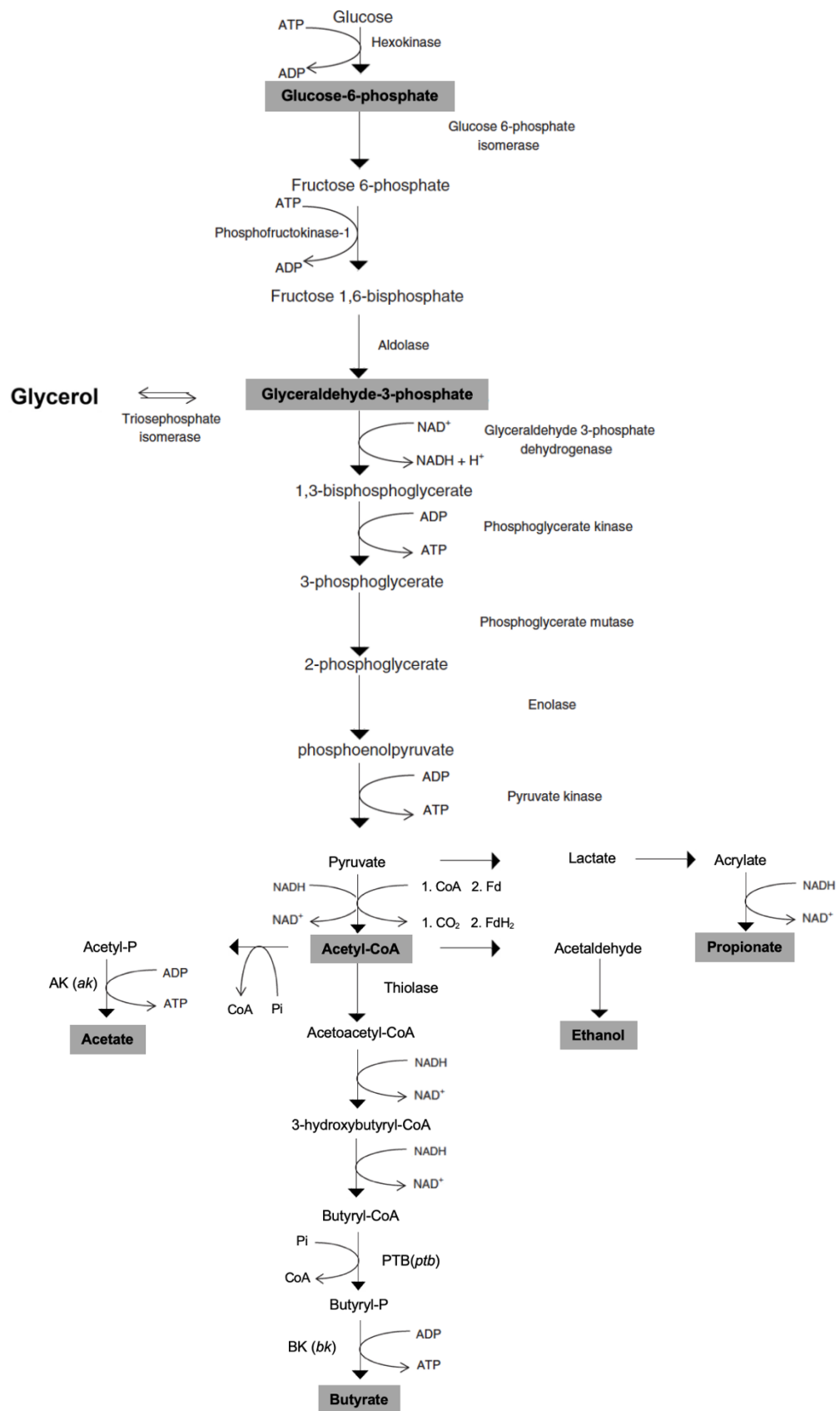


Figure 42. Potential metabolic pathway for glucose and glycerol metabolism; glycolytic pathway. Adapted from [235-237]

SECTION 5

Future work

12 CHAPTER 12: DISCUSSION (Summary, conclusions, future work)

This thesis aimed to demonstrate the augmentation *in vivo* of target volatile organic compounds using judiciously selected nutritional metabolic substrates. This was achieved by (i) a thorough literature review providing a proof of concept of augmentation of short chain fatty acids by bacterial stimulation with nutritional substrates in an industrial context (ii) identification of bacteria related to oesophagogastric cancer through 16S rRNA gene sequencing and Shotgun metagenomics (iii) *in vitro* augmentation of liquid cultures by headspace VOC analysis (iv) augmentation *in vivo* of target VOCs within exhaled breath (v) establishing the source of VOCs detected within exhaled breath by compartmentalising the lungs and gastrointestinal tract intra-operatively and (vi) suggesting the potential role of the oral microbiome in contributing to VOCs detected in exhaled breath.

This work has identified that *Lactobacillus*, *Streptococcus*, *Veillonella* and *Prevotella* are associated with gastric adenocarcinoma and *Streptococcus*, *Veillonella* and *Prevotella* to oesophageal adenocarcinoma. *Lactobacillus* in particular, has been a consistent finding in literature linked to gastric cancer (Table 35).

Table 355. Summary of bacteria genera identified consistently between literature and original microbial analysis in gastric adenocarcinoma.

Gastric adenocarcinoma				
Bacterial genera	Published literature	References	16S rRNA gene sequencing	Shotgun metagenomics
<i>Lactobacillus</i>	Present	[57, 60, 62, 65, 66, 68, 75-82]	Present	Present
<i>Streptococcus</i>	Present	[57, 60, 63, 67, 69, 73, 75-77, 81]		Present
<i>Prevotella</i>	Present	[38, 63, 65, 69, 73, 76, 77, 81, 83]	Present	Present
<i>Veillonella</i>	Present	[38, 57, 65, 69, 71, 76, 77]		Present
<i>Escherichia</i>	Present	[62, 71, 76, 80, 82]		Present
<i>Aggregatibacteria</i>	Present	[38, 48, 69]		Present
<i>Granulicatella</i>	Present	[69, 80, 82]		Present
<i>Haemophilus</i>	Present	[57, 69, 74]		Present
<i>Propionibacterium</i>	Present	[70, 75, 78]	Present	

Novel findings of significantly enriched *Actinomyces* and *Rothia* isolated from saliva of patients with oesophagogastric cancer was reported. *Rothia* is an aerobic microbe linked to pathological oral disease, and other cancer types such as colorectal cancer. The enrichment of *Actinomyces* and *Rothia* adds to the current literature pool where *Fusobacterium nucleatum* has been linked with oesophageal cancer saliva.

This work has established the importance of the oral cavity in contributing to VOCs detected in exhaled breath. Butyric acid is a potential breath biomarker associated with oesophagogastric cancer demonstrated in a recent multi-centre study conducted by this group. Stimulating liquid aerobic saliva cultures generated an increase in butyric acid, suggesting for the first time that butyric acid originates from the oral cavity. The oral microbiome is specific to cancer and reflects a change in the microbiome

across the entire upper gastrointestinal tract. Sun *et al* demonstrated a sensitivity of 97% of cancer detection using the oral cavity microbiome only. Previous work by the group have demonstrated butyric acid in the headspace directly over the tumour. Therefore, there may be a contribution from multiple anatomical compartments. The intraoperative study showed a significant increase in fatty acids within the luminal tract rather than systemically within the lung, supporting the notion that VOCs detected in exhaled breath are multi-compartmental.

This thesis demonstrates the proof of principle that target nutritional metabolites can be utilised by bacteria to upregulate volatile compounds, namely short chain fatty acids. However, the responsible metabolic pathway is yet to be investigated and confirmed. Metabolic pathways responsible for the conversion of nutritional substrates to volatile compounds have been postulated based on literature knowledge and using pathway prediction programmes. Specific pathways must be fully defined to allow manipulation of the pathways and ultimately alter disease pathways. Stable isotope tracing with flux analysis will accurately quantify the proportion of metabolites derived from the labelled substrate, including calculations of the intermediaries and end metabolites. Uniformly labelled ^{13}C glucose will be selected as it covers a range of pathways potentially involved in glucose metabolism. Labelling of carbons 1 and 2 will flux through the oxidative and non-oxidative pentose phosphate pathway to generate lactate M+1 metabolite. Lactate M+2 metabolite can be detected by glycolysis. A ratio of lactate M+1 to lactate M+2 can give the relative contributions of the pentose phosphate and glycolytic pathways. Uniformly labelled glucose is rapidly metabolised via the glycolytic pathway with a higher flux. Intermediaries such as glucose-6-phosphate M+3 and fructose bisphosphatase can be identified to confirm involvement of this pathway. Finally, labelling of carbon 3 in a glucose tracer will produce labelled metabolites from the pyruvate to TCA cycle. The use of multiple tracers and therefore multiple metabolite readouts can enhance information output. Flux analysis is commonly employed as it end metabolite levels can quantify fluxes, especially as glycolysis has a high yield. Multiple time point measurements are suggested during this dynamic phase with analysis of intracellular metabolites and their rates of increase. However, once steady state is achieved, tracers can be used to quantify the contribution of specific labelled nutrients to target end metabolites. This experiment is key in verifying the metabolic pathways involved, and if there is a contribution from multiple pathways / metabolites are a combination of multiple pathways.

This work used batch cultures, with close attention to replicating the culture conditions as accurately as possible. It is a limitation that there may have been some differences between batch cultures generated, in view of some headspace VOCs not been consistent among the samples. There are other possibilities such as failure of the headspace extraction design/system. One way to address this challenge is the use of a bioreactor to simulate the oesophagogastric environment in a dynamic fashion where parameters such as temperature, pH, oxygenation, fresh cell media and nutrient stimulant input can be tightly controlled. This experimental set-up with a 500mL bioreactor would be ideal as a pH of 5, microaerophilia (5% oxygen), physiological temperature of 37 degrees with a thermal water jacket, continuous and consistent stirring of the media will mimic the oesophagogastric environment. Cell

media used as the fuel for liquid culture growth can be tapped off to calculate glucose and electrolyte concentrations to maintain an environment suitable for bacterial growth. Headspace VOCs can be extracted at multiple times in this dynamic environment. Input of multiple stable isotope tracers for all components of the oral stimulant drink, and their contribution to the end metabolite pool will be valuable to know which substrates and pathways are preferred in this experimental model. This can provide a reliable, consistent, and reproducible method dynamic VOC headspace extraction.

This work suggests the source of VOC production and augmentation is the cancer-specific bacteria. However, the question remains about the involvement of the cancer tissue and its increased metabolic activity, a known hallmark of cancer. To investigate the VOC production from cancer tissue (i) requires an adequate blood supply and therefore resection specimens would not be suitable, and (ii) targeted headspace analysis which would not be suitable in an intraoperative setting with confounding factors. A challenging but robust solution is the development of cancer organoids. Organoids are three-dimensional multicellular structures cultivated from pluripotent stem cells in an *in vitro* environment to anatomically, physiologically and metabolically mimic the representative organ. Organoid development has been explored in various organs recapitulating physiological activity specific to the organ, *in vitro* and with implantation into mice. Examples include how brain organoids have demonstrated neuronal activity, lung organoids have developed branching airways and alveolae, and endometrial organoids have demonstrated features of early pregnancy in response to hormone treatment. Gastrointestinal organoids have been extensively studied since 2009 when organoids were first described. Colorectal organoids, both benign inflammatory bowel disease based and malignant have been established. Stem cells located at the bottom of intestinal crypts have self-organising and regenerating capabilities from single cells creating multicellular structures. The Lgr5+ stem cells have demonstrated self-renewing properties with the ability to differentiate into all epithelial cell types. A single crypt contains up to six Lgr5+ cells. Development of organoids of the upper gastrointestinal tract are still in its infancy. An optimal growth medium for human *ex vivo* tissue cell culture needs to be sought. A range of tissue culture mediums exist, however, the optimal environment including altered pH levels for the upper gastrointestinal tract need to be considered. Commonly, this media contains energy sources supporting rapid cell division and cell growth, peptides, and antibiotics such as streptomycin sulphate and gentamicin sulphate. Headspace VOC analysis of cancer organoids can determine cancer related biomarkers. This experimental set-up is important to answer the question of whether VOCs may be originating from the highly vascular and highly metabolic tumour tissue.

Successful establishments of upper gastrointestinal organoids present opportunities for disease modelling and targeted diagnostics and therapeutics. In this context, co-culturing of cancer-specific bacteria and cancer organoids can provide detailed information about the oncological-microbiological axis. It is unknown whether the change in microbiome causes cancer, or whether the development of cancer alters the microbiome. This is the most challenging experiment methodologically as optimal conditions needs to be defined for both organoids and bacteria to survive and function as expected.

Further work spans various science interests including cellular biology, microbiology, biochemistry, and analytical chemistry. We envisage that these *in vitro* and *ex vivo* experiments will translate to *in vivo* clinical studies. The clinical development study will inform a multi-centre national study evaluating the diagnostic performance of an oral stimulant drink to define the parameters of augmentation further. The following step would be an external validation study. The relationship between the oral microbiome and the tissue microbiome in cancer needs to be explored further. This thesis suggests for the first time that the oral cavity is important in contributing VOCs that are unique to oesophagogastric cancer. However no direct link has been proposed between the oral and tumour tissue microbiome.

Discovering a unique VOC profile associated directly with a single or mixed cancer specific bacteria can pave the way for diagnostic strategies which forms the mainstay of this thesis. However, this presents the opportunity for the microbiome to be exploited and manipulated with the use of antibiotics to alter the course of the disease process/pathophysiological process.

Diagnostic volatile biomarkers detected within exhaled breath in oesophagogastric cancer patients may be multi-compartmental. Bacteria being the source of VOC production is crucial this thesis, and supporting literature, demonstrates the cancer microbiome is markedly different to that of the healthy population. There is wide scope to investigate alterations in the human ecological system and its role in identifying cancer-related biomarkers.

REFERENCES

1. World Health Organisation Global Status report. <https://www.who.int/news-room/fact-sheets/detail/cancer>. p. [Accessed on 09.11.2019].
2. Global Cancer Observatory. <https://gco.iarc.fr/today/data/factsheets/cancers/>. p. [Accessed on 09.11.2019].
3. National Oesophagogastric Cancer Audit 2019 <https://www.nogca.org.uk> p. [Accessed on 09.11.2019].
4. Wong MCS, H.W., Whiteman DC, Jiang JY, Qiao Y, Fung FDH, Wang HHX, Chiu PWY, Ng EKW, Wu JCY, Yu J, Chan FKL, Sung JY, *Global Incidence and mortality of oesophageal cancer and their correlation with socioeconomic indicators temporal patterns and trends in 41 countries*. Sci Rep, 2018. **8**: p. 4522.
5. World Cancer Research Fund. <https://www.wcrf.org/dietandcancer/stomach-cancer>. p. [Accessed on 09.11.2019].
6. Cancer Research UK <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer>. p. [Accessed on 10/06/2019].
7. Wang VS, H.J., Sepulveda JA, Mauer R, Poneros JM, *Low prevalence of submucosal invasive carcinoma at esophagectomy for high-grade dysplasia or intramucosal adenocarcinoma in Barrett's esophagus: a 20-year experience*. Gastrointest Endosc, 2009. **69**: p. 777-783.
8. Rolff HC, S.L., Rosenberg J, *Clinical findings confirm national guidelines regarding primary gastroscopy for upper gastrointestinal symptoms*. Dan Med Bull, 2011. **58**: p. A4363.
9. Vakil N, M.P., Fennerty MB, Talley NJ, *Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: systematic review and meta-analysis*. Gastroenterology, 2006. **131**: p. 390-401.
10. Creemers A, E.E., Hooijer GJ, Stap L, Jibodh-Mulder RA, Gisbertz SS, van Berge Henegouwen MI, van Montfoort ML, Hulshof MCCM, Krishnadath KK, van Oijen MGH, Bijlsma MF, Meijer SL, van Laarhoven HWM, *The dynamics of HER2 status in esophageal adenocarcinoma*. Oncotarget, 2018. **9**: p. 26787-26799.
11. Walsh EM, K.R., *Single agent anti PD-1 inhibitors in esophageal cancer-a first step in a new therapeutic direction*. J Thorac Dis, 2018. **10**: p. 1308-1313.
12. Stroun M, A.P., Lyautey J, Lederrey C, Maurice PA, *Isolation and characterization of DNA from the plasma of cancer patients*. Eur J Cancer Clin Oncol, 1987. **23**: p. 707-712.
13. Leon SA, S.B., Sklaroff DM, Yaros MJ, *Free DNA in the serum of cancer patients and the effect of therapy*. Cancer Res, 1977. **37**: p. 646-650.
14. Wu TL, Z.D., Chia JH, Tsao K-, Sun CF, Wu JT, *Cell-free DNA: measurement in various carcinomas and establishment of normal reference range*. Clin Chim Acta, 2002. **321**: p. 77-87.
15. Testa U, C.G., Pelosi E, *Esophageal Cancer: Genomic and Molecular Characterization, Stem Cell Compartment and Clonal Evolution*. Medicines (Basel), 2017. **4**: p. 67.
16. Xu YW, C.H., Guo HP, Yang SH, Luo YH, Liu CT, Huang XY, Tang XM, Hong CQ, Li EM, Xu LY, Peng YH, *Combined detection of serum autoantibodies as diagnostic*

- biomarkers in esophagogastric junction adenocarcinoma*. Gastric Cancer, 2019. **22**: p. 546-557.
17. Matsuoka T, Y.M., *Biomarkers of gastric cancer: Current topics and future perspective*. World J Gastroenterol, 2018. **24**: p. 2818-2832.
 18. Qu Y, D.S., Hou P, *Gene methylation in gastric cancer*. Clin Chim Acta, 2013. **424**: p. 53-65.
 19. Sapari NS, L.M., Vaithilingam A, Soong R, *Clinical potential of DNA methylation in gastric cancer: a meta-analysis*. PLoS One, 2012. **7**: p. e36275.
 20. Ding Y, L.H., Du Y, *Clinicopathological significance of CHFR promoter methylation in gastric cancer: a meta-analysis*. Oncotarget, 2017. **9**: p. 10083-10090.
 21. Gao Y, Z.K., Xi H, Cai A, Wu X, Cui J, Li J, Qiao Z, Wei B, Chen L, *Diagnostic and prognostic value of circulating tumor DNA in gastric cancer: a meta-analysis*. Oncotarget, 2017. **8**: p. 6330-6340.
 22. Baniak N, S.J., Ahmed S, Kanthan SC, Kanthan R, *Gastric biomarkers: a global review*. World J Surg Oncol, 2016. **14**: p. 212.
 23. Shah AK, S.N., Barbour AP, Hill MM, *Early diagnostic biomarkers for esophageal adenocarcinoma – the current state of play*. Cancer Epidemiol Biomarkers Prev, 2013. **22**: p. 1185-1209.
 24. Liu K, Z.T., Wang J, Chen Y, Zhang R, Lan X, Que J, *Etiology, cancer stem cells and potential diagnostic biomarkers for esophageal cancer*. Cancer Lett, 2019. **458**: p. 21-28.
 25. Hardikar S, O.L., Song X, Wilson AM, Montine TJ, Kratz M, Anderson GL, Blount PL, Reid BJ, White E, Vaughan TL, *Inflammation and oxidative stress markers and esophageal adenocarcinoma incidence in a Barrett's esophagus cohort*. Cancer Epidemiol Biomarkers Prev, 2014. **23**: p. 2393-2403.
 26. Siahpush SH, V.T., Lampe JN, Freeman R, Lewis S, Odze RD, Blount PL, Ayub K, Rabinovitch PS, Reid BJ, Chen C, *Longitudinal study of insulin-like growth factor, insulin-like growth factor binding protein-3, and their polymorphisms: risk of neoplastic progression in Barrett's esophagus*. Cancer Epidemiol Biomarkers Prev, 2007. **16**: p. 2387-2395.
 27. Keeley BR, I.F., Pourshams A, Poustchi H, Pak JS, Brennan P, Khademi H, Genden EM, Abnet CC, Dawsey SM, Boffetta P, Malekzadeh R, Sikora AG, *Prediagnostic serum levels of inflammatory biomarkers are correlated with future development of lung and esophageal cancer*. Cancer Sci, 2014. **105**: p. 1205-1211.
 28. Tu H, S.L., Dong X, Gong Y, Xu Q, Jing J, Bostick RM, Wu X, Yuan Y, *A serological biopsy using five stomach-specific circulating biomarkers for gastric cancer risk assessment: A multi-phase study*. Am J Gastroenterol, 2017. **112**: p. 704-715.
 29. Leja M, Y.W., Camargo MC, Saito H, *Implementation of gastric cancer screening - the global experience*. Best Pract Res Clin Gastroenterol, 2014. **28**: p. 1093-1106.
 30. Lee HS, J.S., Nomura S, Seto Y, Kwon YH, Nam SY, Ishibashi Y, Ohtsu H, Ohmoto Y, Yang HM, *Screening biomarker as an alternative to endoscopy for the detection of early gastric cancer: The combination of serum trefoil factor family 3 and pepsinogen*. Gastroenterol Res Pract, 2018. **2018**: p. 1024074.
 31. Kumar S, H.J., Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, Lovat LB, Španěl P, Smith D, Hanna GB, *Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma*. Ann Surg, 2015. **262**: p. 981-990.

32. Markar SR, W.T., Antonowicz S, Chin ST, Romano A, Nikolic K, Evans B, Cunningham D, Mughal M, Lagergren J, Hanna GB, *Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer*. JAMA Oncol, 2018. **4**: p. 970-976.
33. Adam ME, F.M., Boshier PR, Chin ST, Lin GP, Romano A, Kumar S, Hanna GB, *Mass-Spectrometry Analysis of Mixed-Breath, Isolated-Bronchial-Breath, and Gastric-Endoluminal-Air Volatile Fatty Acids in Esophagogastric Cancer*. Anal Chem, 2019. **91**: p. 3740-3746.
34. Ursell LK, M.J., Parfrey LW, Knight R, *Defining the human microbiome*. Nutr Rev, 2012. **70**: p. S38-44.
35. Kelly D, Y.L., Pei Z, *A review of the oesophageal microbiome in health and disease*. The Human Microbiome: Methods in Microbiology, 2017. **44**: p. 19-35.
36. Johansson I, W.E., Kaveh B, Lif Holgersson P, Tanner AC, *The microbiome in populations with a low and high prevalence of caries*. J Dent Res, 2016. **95**: p. 80-86.
37. Weng MT, C.Y., Wei PY, Chiang CW, Fang HL, Wei SC, *Microbiota and gastrointestinal cancer*. J Formos Med Assoc, 2019. **118**: p. S32-41.
38. Sun JH, L.X., Yin J, Li YH, Hou BX, Zhang Z, *A screening method for gastric cancer by oral microbiome detection*. Oncol Rep, 2018. **39**: p. 2217-2224.
39. Marchesi JR, A.D., Fava F, Hermes GD, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, Thomas LV, Zoetendal EG, Hart A, *The gut microbiota and host health: a new clinical frontier*. Gut, 2016. **65**: p. 330-339.
40. Pimentel M, S.R., Long MD, Rao SSC, *ACG Clinical Guideline: Small Intestinal Bacterial Overgrowth*. Am J Gastroenterol, 2020. **115**: p. 165-178.
41. Ianiro G, R.E., Thomas AM, Schinzari G, Masucci L, Quaranta G, Settanni CR, Lopetuso LR, Armanini F, Blanco-Miguez A, Asnicar F, Consolandi C, Iacovelli R, Sanguinetti M, Tortora G, Gasbarrini A, Segata N, Cammarota G, *Faecal microbiota transplantation for the treatment of diarrhoea induced by tyrosine-kinase inhibitors in patients with metastatic renal cell carcinoma*. Nat Commun, 2020. **11**: p. 4333.
42. Baba Y, I.M., Yoshida N, Watanabe M, Baba H, *Review of the gut microbiome and oesophageal cancer: Pathogenesis and potential clinical implications*. Ann Gastroenterol Surg, 2017. **1**: p. 99-104.
43. Bik EM, E.P., Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, Relman DA, *Molecular analysis of the bacterial microbiota in the human stomach*. Proc Natl Acad Sci U S A, 2006. **103**: p. 732-737.
44. Dicksved J, L.M., Rosenquist M, Enroth H, Jansson JK, Engstrand L, *Molecular characterization of the stomach microbiota in patients with gastric cancer and in controls*. J Med Microbiol, 2009. **58**: p. 509-516.
45. Pereira-Marques J, F.R., Pinto-Ribeiro I, Figueiredo C, *Helicobacter pylori infection, the gastric microbiome and gastric cancer*. Adv Exp Med Biol, 2019. **1149**: p. 195-210.
46. Coker OO, D.Z., Nie Y, Zhao G, Cao L, Nakatsu G, Wu WK, Wong SH, Chen Z, Sung JJY, Yu J, *Mucosal microbiome dysbiosis in gastric carcinogenesis*. Gut, 2018. **67**: p. 1024-1032.
47. Peters BA, W.J., Pei Z, Yang L, Purdue MP, Freedman ND, Jacobs EJ, Gapstur SM, Hayes RB, Ahn J, *Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers*. Cancer Res, 2017. **77**: p. 6777-6787.

48. Hu YL, P.W., Huang Y, Zhang Y, Zhang CJ, *The Gastric Microbiome Is Perturbed in Advanced Gastric Adenocarcinoma Identified Through Shotgun Metagenomics*. Front Cell Infect Microbiol, 2018. **8**: p. 433.
49. Gaude E, N.M., Patassini S, Boschmans J, Allsworth M, Boyle B, van der Schee MP, *Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes*. J Breath Res, 2019. **13**: p. 032001.
50. Humans, I.W.G.o.t.E.o.C.R.t., *Schistosomes, Liver Flukes and Helicobacter pylori*. IARC Monogr Eval Carcinog Risks Hum. , 1994. **61**: p. 1-241.
51. Fan X, A.A., Wu J, Peters BA, Jacobs EJ, Gapstur SM, Purdue MP, Abnet CC, Stolzenberg-Solomon R, Miller G, Ravel J, Hayes RB, Ahn J, *Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study*. Gut, 2018. **67**: p. 120-127.
52. Flemer B, W.R., Barrett MP, Cisek K, Das A, Jeffery IB, Hurley E, O'Riordain M, Shanahan F, O'Toole PW, *The oral microbiota in colorectal cancer is distinctive and predictive*. Gut, 2018. **67**: p. 1454-1463.
53. Hosgood HD, C.Q., Hua X, Long J, Shi J, Wan Y, Yang Y, Abnet C, Bassig BA, Hu W, Ji BT, Klugman M, Xiang Y, Gao YT, Wong JY, Zheng W, Rothman N, Shu XO, Lan Q, *Variation in oral microbiome is associated with future risk of lung cancer among never-smokers*. Thorax, 2021. **76**: p. 256-263.
54. Reddy RM, W.W., Barnett S, Heiden BT, Orringer MB, Lin J, Chang AC, Carrott PW, Lynch WR, Beer DG, Fenno JC, Kapila Y, *Increased Variance in Oral and Gastric Microbiome Correlates With Esophagectomy Anastomotic Leak*. Ann Thorac Surg, 2018. **105**: p. 865-870.
55. Yuda M, Y.K., Okamura A, Hayami M, Fukudome I, Toihata T, Imamura Y, Mine S, Ishizuka N, Watanabe M, *Influence of Preoperative Oropharyngeal Microflora on the Occurrence of Postoperative Pneumonia and Survival in Patients Undergoing Esophagectomy for Esophageal Cancer*. Ann Surg, 2020. **272**: p. 1035-1043.
56. Mirzayi C, R.A., Zohra F, Elsafoury S, Kasselmann L, van de Wijgert J, Segata N, Beghini F, Eckenrode K, Dowd J, Jones, HE, Waldron L., *Strengthening The Organizing and Reporting of Microbiome Studies (STORMS)*. bioRxiv, 2020.
57. Sjostedt S, H.A., Kager L, Nord CE, *Microbial colonization of the oropharynx, esophagus and stomach in patients with gastric diseases*. Eur J Clin Microbiol, 1985. **4**: p. 49-51.
58. Seo I, J.B., Suh S, Suh M, Baek W, *Microbial profile of the stomach: Comparison between normal mucosa and cancer tissue in the same patient*. J Bacteriol Virol, 2014. **44**: p. 162-169.
59. Aviles-Jimenez F, V.-J.F., Medrano-Guzman R, Mantilla A, Torres J, *Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer*. Sci Rep, 2014. **4**: p. 4202.
60. Eun CS, K.B., Han DS, Kim SY, Kim KM, Choi BY, Song KS, Kim YS, Kim JF, *Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods*. Helicobacter, 2014. **19**: p. 407-416.
61. Hu J, H.S., Chen Y, Ji Z, *Variations of tongue coating microbiota in patients with gastric cancer*. Biomed Res Int, 2015: p. 173729.

62. Wang L, Z.J., Xin Y, Geng C, Tian Z, Yu X, Dong Q, *Bacterial overgrowth and diversification of microbiota in gastric cancer*. Eur J Gastroenterol Hepatol, 2016. **28**: p. 261-266.
63. Yu G, T.J., Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, Wang L, Wang C, Ding T, Ravel J, Taylor PR, Abnet CC, Goldstein AM, *Molecular Characterization of the Human Stomach Microbiota in Gastric Cancer Patients*. Front Cell Infect Microbiol, 2017. **7**: p. 302.
64. Li TH, Q.Y., Sham PC, Lau KS, Chu KM, Leung WK, *Alterations in Gastric Microbiota After H. Pylori Eradication and in Different Histological Stages of Gastric Carcinogenesis*. Sci Rep, 2017. **7**: p. 44935.
65. Castaño-Rodríguez N, G.K., Fock KM, Mitchell HM, Kaakoush NO, *Dysbiosis of the microbiome in gastric carcinogenesis*. Sci Rep, 2017. **7**: p. 15957.
66. Ferreira RM, P.-M.J., Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, Figueiredo C, *Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota*. Gut, 2018. **67**: p. 226-236.
67. Wu J, X.S., Xiang C, Cao Q, Li Q, Huang J, Shi L, Zhang J, Zhan Z, *Tongue Coating Microbiota Community and Risk Effect on Gastric Cancer*. J Cancer, 2018. **9**: p. 4039-4048.
68. Hsieh YY, T.S., Pan HY, Yen CW, Xu HW, Lin YJ, Deng YF, Hsu WT, Wu CS, Li C, *Increased Abundance of Clostridium and Fusobacterium in Gastric Microbiota of Patients with Gastric Cancer in Taiwan*. Sci Rep, 2018. **8**: p. 158.
69. Shao D, V.E., Liu A, Qin J, Chen W, Abnet CC, Wei W, *Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China*. Cancer, 2019. **125**: p. 3993-4002.
70. Gunathilake MN, L.J., Choi IJ, Kim YI, Ahn Y, Park C, Kim J, *Association between the relative abundance of gastric microbiota and the risk of gastric cancer: a case-control study*. Sci Rep, 2019. **9**: p. 13589.
71. Liang W, Y.Y., Wang H, Wang H, Yu X, Lu Y, Shen S, Teng L, *Gut microbiota shifts in patients with gastric cancer in perioperative period*. Medicine (Baltimore), 2019. **98**: p. e16626.
72. Kageyama S, T.T., Takeuchi K, Asakawa M, Matsumi R, Furuta M, Shibata Y, Nagai K, Ikebe M, Morita M, Masuda M, Toh Y, Kiyohara Y, Ninomiya T, Yamashita Y, *Characteristics of the Salivary Microbiota in Patients With Various Digestive Tract Cancers*. Front Microbiol, 2019. **10**: p. 1780.
73. Chen XH, W.A., Chu AN, Gong YH, Yuan Y, *Mucosa-Associated Microbiota in Gastric Cancer Tissues Compared With Non-cancer Tissues*. Front Microbiol, 2019. **10**: p. 1261.
74. Dong Z, C.B., Pan H, Wang D, Liu M, Yang Y, Zou M, Yang J, Xiao K, Zhao R, Zheng X, Zhang L, Zhang Y, *Detection of Microbial 16S rRNA Gene in the Serum of Patients With Gastric Cancer*. Front Oncol, 2019. **9**: p. 608.
75. Liu X, S.L., Liu X, Ji F, Mei Y, Cheng Y, Liu F, Yan C, Li L, Ling Z, *Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer*. EBioMedicine, 2019. **40**: p. 336-348.
76. Qi YF, S.J., Ren LF, Cao XL, Dong JH, Tao K, Guan XM, Cui YN, Su W, *Intestinal Microbiota Is Altered in Patients with Gastric Cancer from Shanxi Province, China*. Dig Dis Sci, 2019. **64**: p. 1193-1203.

77. Wang Z, G.X., Zeng R, Wu Q, Sun H, Wu W, Zhang X, Sun G, Yan B, Wu L, Ren R, Guo M, Peng L, Yang Y, *Changes of the Gastric Mucosal Microbiome Associated With Histological Stages of Gastric Carcinogenesis*. Front Microbiol, 2020. **11**: p. 997.
78. Wang L, X.Y., Zhou J, Tian Z, Liu C, Yu X, Meng X, Jiang W, Zhao S, Dong Q, *Gastric Mucosa-Associated Microbial Signatures of Early Gastric Cancer*. Front Microbiol, 2020. **11**: p. 1548.
79. Spiegelhauer MR, K.J., Johannesen TB, Urba M, Skieceviciene J, Jonaitis L, Frandsen TH, Kupcinskas L, Fuursted K, Andersen LP, *Transient and Persistent Gastric Microbiome: Adherence of Bacteria in Gastric Cancer and Dyspeptic Patient Biopsies after Washing*. J Clin Med, 2020. **9**: p. 1882.
80. Gantuya B, E.S.H., Matsumoto T, Ajami NJ, Uchida T, Oyuntsetseg K, Bolor D, Yamaoka Y, *Gastric mucosal microbiota in a Mongolian population with gastric cancer and precursor conditions*. Aliment Pharmacol Ther, 2020. **51**: p. 770-780.
81. Wu ZF, Z.K., Wu GN, Jin ZJ, Xiang CJ, Xu S, Wang YH, Wu XY, Chen C, Xu Z, Li WS, Yao XQ, Zhang JF, Liu FK, *A Comparison of Tumor-Associated and Non-Tumor-Associated Gastric Microbiota in Gastric Cancer Patients*. Dig Dis Sci, 2020.
82. Xu S, X.C., Wu J, Teng Y, Wu Z, Wang R, Lu B, Zhan Z, Wu H, Zhang J, *Tongue Coating Bacteria as a Potential Stable Biomarker for Gastric Cancer Independent of Lifestyle*. Dig Dis Sci, 2020.
83. Dang YN, D.Y., Mu YZ, Yan J, Lu M, Zhu YL, Zhang GX, *Identification of gastric microbiota biomarker for gastric cancer*. Chin Med J 2020. **133**: p. 2765-2767.
84. Chen X, W.B., Lu M, Cheng H, Yuan Z, Yang Y, Jin L, Ye W, *Oral microbiota and risk for esophageal squamous cell carcinoma in a high-risk area of China*. PLoS One, 2015. **10**: p. e0143603.
85. Nasrollahzadeh D, M.R., Ploner A, Shakeri R, Sotoudeh M, Fahimi S, Nasser-Moghaddam S, Kamangar F, Abnet CC, Winckler B, Islami F, Boffetta P, Brennan P, Dawsey SM, Ye W, *Variations of gastric corpus microbiota are associated with early esophageal squamous cell carcinoma and squamous dysplasia*. Sci Rep, 2015. **5**: p. 8820.
86. Yamamura K, B.Y., Nakagawa S, Mima K, Miyake K, Nakamura K, Sawayama H, Kinoshita K, Ishimoto T, Iwatsuki M, Sakamoto Y, Yamashita Y, Yoshida N, Watanabe M, Baba H, *Human Microbiome Fusobacterium Nucleatum in Esophageal Cancer Tissue Is Associated with Prognosis*. Clin Cancer Res, 2016. **22**: p. 5574-5581.
87. Elliott DRF, W.A., O'Donovan M, Parkhill J, Fitzgerald RC, *A non-endoscopic device to sample the oesophageal microbiota: a case-control study*. Lancet Gastroenterol Hepatol, 2017. **2**: p. 32-42.
88. Wang Q, R.Y., Guo X, Liu N, Liu S, Wen P, Li S, Li Y, *Oral Microbiome in Patients with Oesophageal Squamous Cell Carcinoma*. Sci Rep, 2019. **9**: p. 19055.
89. Yamamura K, I.D., Kandimalla R, Sonohara F, Baba Y, Yoshida N, Kodera Y, Baba H, Goel A, *Intratumoral Fusobacterium Nucleatum Levels Predict Therapeutic Response to Neoadjuvant Chemotherapy in Esophageal Squamous Cell Carcinoma*. Clin Cancer Res, 2019. **25**: p. 6170-6179.
90. Xu L, L.Y., Sun S, Yue J, *Decrease of oral microbial diversity might correlate with radiation esophagitis in patients with esophageal cancer undergoing chemoradiation: A pilot study*. Prec Radiat Oncol, 2020. **4**: p. 81-88.
91. Peter S, P.A., VanDerPol W, Wilcox CM, Kyanam Kabir Baig KR, Morrow C, Izard J, Mannon PJ, *Mucosa-Associated Microbiota in Barrett's Esophagus, Dysplasia, and*

- Esophageal Adenocarcinoma Differ Similarly Compared With Healthy Controls*. Clin Transl Gastroenterol, 2020. **11**: p. e00199.
92. Li D, H.R., Hou G, Ming W, Fan T, Chen L, Zhang L, Jiang W, Wang W, Lu Z, Feng H, Geng Q, *Characterization of the Esophageal Microbiota and Prediction of the Metabolic Pathways Involved in Esophageal Cancer*. Front Cell Infect Microbiol, 2020. **10**: p. 268.
 93. Zhao Q, Y.T., Yan Y, Zhang Y, Li Z, Wang Y, Yang J, Xia Y, Xiao H, Han H, Zhang C, Xue W, Zhao H, Chen H, Wang B, *Alterations of Oral Microbiota in Chinese Patients With Esophageal Cancer*. Front Cell Infect Microbiol, 2020. **10**: p. 541144.
 94. Zhou J, S.P., Qiu Z, Harman DG, Teoh WC, Al-Sohaily S, Liem H, Turner I, Ho V, *Distinct Microbiota Dysbiosis in Patients with Non-Erosive Reflux Disease and Esophageal Adenocarcinoma*. J Clin Med, 2020. **9**: p. 2162.
 95. Lopetuso LR, S.M., Pecere S, Ponziani FR, Boskoski I, Larghi A, Quaranta G, Masucci L, Ianaro G, Camboni T, Gasbarrini A, Costamagna G, Consolandi C, Cammarota G, *Esophageal microbiome signature in patients with Barrett's esophagus and esophageal adenocarcinoma*. PLoS One, 2020. **15**: p. e0231789.
 96. Kawasaki M, I.Y., Ikeda E, Takahashi M, Tanaka D, Nakajima Y, Arakawa S, Izumi Y, Miyake S, *Oral infectious bacteria in dental plaque and saliva as risk factors in patients with esophageal cancer*. Cancer, 2021. **127**: p. 512-519.
 97. Wells GA, S.B., O-Connell D, Peterson J, Welch V, Losos M, Tugwell P, *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
 98. Arnold M, S.I., Ferlay J, Forman D, *Global incidence of oesophageal cancer by histological subtype in 2012*. Gut, 2015. **64**: p. 381-387.
 99. Bhat JV, B.H., *Clostridium lacto-acetophilum nov. spec. and the role of acetic acid in the butyric acid fermentation of lactate*. J Bacteriol, 1947. **54**: p. 381-391.
 100. AT, J., *Isolation of a bacterium, producing propionic acid, from the rumen of sheep*. J Gen Microbiol, 1951. **5**: p. 317-325.
 101. Namdari H, C.V., *Glucose-mediated catabolite repression of the tricarboxylic acid cycle as an explanation for increased acetic acid production in suicidal Aeromonas strains*. J Bacteriol, 1990. **172**: p. 4721-4724.
 102. Lawford HG, R.J., *Fermentation of biomass-derived glucuronic acid by pet expressing recombinants of E. coli B*. Appl Biochem Biotechnol, 1997. **63-65**: p. 221-241.
 103. Meurice G, B.A., Deborde C, Boyaval P, *Determination of metabolic fluxes during glucose catabolism in Propionibacterium freudenreichii subsp. shermanii*. Meded Rijksuniv Gent Fak Landbouwkde Toegep Biol Wet, 2001. **66**: p. 219-225.
 104. Jiang L, W.J., Liang S, Wang X, Cen P, Xu Z, *Production of butyric acid from glucose and xylose with immobilized cells of Clostridium tyrobutyricum in a fibrous-bed bioreactor*. Appl Biochem Biotechnol, 2010. **160**: p. 350-359.
 105. He F, Q.S., Yang Z, Bai X, Suo Y, Wang J, *Butyric acid production from spent coffee grounds by engineered Clostridium tyrobutyricum overexpressing galactose catabolism genes*. Bioresour Technol, 2020. **304**: p. 122977.
 106. Lee KM, C.O., Kim KY, Woo HM, Kim Y, Han SO, Sang BI, Um Y, *Butyric acid production from red algae by a newly isolated Clostridium sp. S1*. Biotechnol Lett, 2015. **37**: p. 1837-1844.

107. Shi Z, H.L., Wu X, Luo L, Xiao K, Cai J, Xu Z, *Long-term production of butyric acid through immobilization of Clostridium tyrobutyricum in a moving fibrous-bed bioreactor (MFBB)*. J Chem Technol Biotechnol, 2013. **89**: p. 1883-1889.
108. Song JH, V.J., Lee CH, Jahng D, *Butyric acid production from brown algae using Clostridium tyrobutyricum ATCC 25755*. Biotechnol Bioprocess Eng, 2011. **16**: p. 42-49.
109. Mitchell RJ, K.J., Jeon BS, Sang BI, *Continuous hydrogen and butyric acid fermentation by immobilized Clostridium tyrobutyricum ATCC 25755: Effects of the glucose concentration and hydraulic retention time*. Bioresour Technol, 2009. **100**: p. 5352-5355.
110. Wu Z, Y.S., *Extractive fermentation for butyric acid production from glucose by Clostridium tyrobutyricum*. Biotechnol Bioeng, 2003. **82**: p. 93-102.
111. Rastogi G, G.R., Bhalla A, Gonzalez R, Bischoff KM, Hughes SR, Kumar S, Sani RK, *Presence of glucose, xylose, and glycerol fermenting bacteria in the deep biosphere of the former Homestake gold mine, South Dakota*. Front Microbiol, 2013. **4**: p. 18.
112. Zhou X, W.S., Lu XH, Liang JP, *Comparison of the effects of high energy carbon heavy ion irradiation and Eucommia ulmoides Oliv. on biosynthesis butyric acid efficiency in Clostridium tyrobutyricum*. Bioresour Technol, 2014. **161**: p. 221-229.
113. Abbad-Andaloussi S, D.C., Raval G, Petitdemange H, *Carbon and electron flow in Clostridium butyricum grown in chemostat culture on glycerol and on glucose*. Microbiology (Reading), 1996. **142**: p. 1149-1158.
114. Evans PJ, W.H., *Effects of extractive fermentation on butyric acid production by Clostridium acetobutylicum*. Appl Microbiol Biotechnol, 1990. **32**: p. 393-397.
115. Michel-Savin D, M.R., Vandecasteele JP, *Control of the selectivity of butyric acid production and improvement of fermentation performance with Clostridium tyrobutyricum*. Appl Microbiol Biotechnol, 1990. **32**: p. 387-392.
116. Forsberg CW, D.L., Gibbins LN, *Metabolism of rhamnose and other sugars by strains of Clostridium acetobutylicum and other Clostridium species*. Can J Microbiol, 1987. **33**: p. 21-26.
117. Turton LJ, D.D., Ganguli LA, *Effect of glucose concentration in the growth medium upon neutral and acidic fermentation end-products of Clostridium bifermentans. Clostridium sporogenes and Peptostreptococcus anaerobius*. J Med Microbiol, 1983. **16**: p. 61-67.
118. SH, C., *Fermentation of 1,2-O-iso-propylidene-D-glucofuranose ("monoacetone glucose") by anaerobic bacteria*. Zentralbl Bakteriell, 1980. **247**: p. 495-501.
119. Sugaya K, T.D., Jones JL, *Production of acetic acid by Clostridium thermoaceticum in batch and continuous fermentations*. Biotechnol Bioeng, 1986. **28**: p. 678-683.
120. Wang G, W.D., *Production of acetic acid by immobilized whole cells of Clostridium thermoaceticum*. Appl Biochem Biotechnol, 1983. **8**: p. 491-503.
121. Concha ML, C.A., Liébana J, Gutiérrez J, Garcia-Mendoza A, *Initial pH as a determining factor of glucose consumption and lactic and acetic acid production in oral streptococci*. Microbios, 1996. **87**: p. 207-216.
122. Arkoudelos JS, N.G., *Comparative studies of the growth of Staphylococcus carnosus with or without glucose*. Lett Appl Microbiol, 1995. **20**: p. 19-24.
123. Robbins JW Jr, T.K., *Optimization of Escherichia coli growth by controlled addition of glucose*. Biotechnol Bioeng, 1989. **34**: p. 1289-1294.

124. Selder L, S.W., Jürgensen N, Lakshmanan A, Zeng AP, *Co-cultures with integrated in situ product removal for lactate-based propionic acid production*. Bioprocess Biosyst Eng, 2020. **43**: p. 1027-1035.
125. Paramithiotis S, S.A., Tsakalidou E, Kalantzopoulos G, *Flour carbohydrate catabolism and metabolite production by sourdough lactic acid bacteria*. World J Microbiol Biotechnol, 2007. **23**: p. 1417-1423.
126. Saha BC, N.L., *Production of mannitol and lactic acid by fermentation with Lactobacillus intermedius NRRL B-3693*. Biotechnol Bioeng, 2003. **82**: p. 864-871.
127. Pérez Chaia A, S.d.S.A., de Ruiz Holgado AP, Oliver G, *Short-chain fatty acids modulate growth of lactobacilli in mixed culture fermentations with propionibacteria*. Int J Food Microbiol, 1995. **26**: p. 365-374.
128. Tyree RW, C.E., Gaddy JL, *The production of propionic acid from sugars by fermentation through lactic acid as an intermediate*. J Chem Technol Biotechnol, 1991. **50**: p. 157-166.
129. Zhang A, S.J., Wang Z, Yang ST, Zhou H, *Effects of carbon dioxide on cell growth and propionic acid production from glycerol and glucose by Propionibacterium acidipropionici*. Bioresour Technol, 2015. **175**: p. 374-381.
130. Stowers CC, C.B., Rodriguez BA, *Development of an industrializable fermentation process for propionic acid production*. J Ind Microbiol Biotechnol, 2014. **41**: p. 837-852.
131. Liu Y, Z.Y., Zhang RB, Zhang F, Zhu J, *Glycerol/glucose co-fermentation: One more proficient process to produce propionic acid by Propionibacterium acidipropionici*. Curr Microbiol, 2011. **62**: p. 152-158.
132. Suwannakham S, Y.S., *Enhanced propionic acid fermentation by Propionibacterium acidipropionici mutant obtained by adaptation in a fibrous-bed bioreactor*. Biotechnol Bioeng, 2005. **91**: p. 325-337.
133. Himmi EH, B.A., Boussaid A, Hassani L, *Propionic acid fermentation of glycerol and glucose by Propionibacterium acidipropionici and Propionibacterium freudenreichii ssp. shermanii*. Appl Microbiol Biotechnol, 2000. **53**: p. 435-440.
134. Ozadali F, G.B., Glatz CE, *Fed-batch fermentation with and without on-line extraction for propionic and acetic acid production by Propionibacterium acidipropionici*. Appl Microbiol Biotechnol, 1996. **44**: p. 710-716.
135. Lewis VP, Y.S., *Propionic acid fermentation by Propionibacterium acidipropionici: effect of growth substrate*. Appl Microbiol Biotechnol, 1992. **37**.
136. Woskow SA, G.B., *Propionic acid production by a propionic acid-tolerant strain of Propionibacterium acidipropionici in batch and semicontinuous fermentation*. Appl Environ Microbiol, 1991. **57**: p. 2821-2828.
137. Feng XH, C.F., Xu H, Wu B, Yao J, Ying HJ, Ouyang PK, *Propionic acid fermentation by Propionibacterium freudenreichii CCTCC M207015 in a multi-point fibrous-bed bioreactor*. Bioprocess Biosyst Eng, 2010. **33**: p. 1077-1085.
138. Frantz JC, M.R., *Growth yields and fermentation balance of Bacteroides fragilis cultured in glucose-enriched medium*. J Bacteriol, 1979. **137**: p. 1263-1270.
139. Wang Z, Y.S., *Propionic acid production in glycerol/glucose co-fermentation by Propionibacterium freudenreichii subsp. shermanii*. Bioresour Technol, 2013. **137**: p. 116-123.
140. Rios-Covian D, A.S., Hernandez-Barranco AM, Alvarez-Buylla JR, Ruas-Madiedo P, Gueimonde M, de los Reyes-Gavilan CG, *Interactions between Bifidobacterium and*

- Bacteroides* species in cofermentations are affected by carbon sources, including exopolysaccharides produced by bifidobacteria. Appl Environ Microbiol, 2013. **79**: p. 7518-7524.
141. Van der Meulen R, A.L., De Vuyst L, *Short fractions of oligofructose are preferentially metabolized by Bifidobacterium animalis DN-173 010*. Appl Environ Microbiol, 2004. **70**: p. 1923-1930.
 142. Liu D, W.S., Xu B, Guo Y, Zhao J, Liu W, Sun Z, Shao C, Wei X, Jiang Z, Wang X, Liu F, Wang J, Huang L, Hu D, He X, Riedel CU, Yuan J, *Proteomics analysis of Bifidobacterium longum NCC2705 growing on glucose, fructose, mannose, xylose, ribose, and galactose*. Proteomics, 2011. **11**: p. 2628-2638.
 143. Huang J, T.W., Zhu S, Du M, *Biosynthesis of butyric acid by Clostridium tyrobutyricum*. Prep Biochem Biotechnol, 2018. **48**: p. 427-434.
 144. Martínez-Anaya MA, L.M., Pilar Macías M, Collar C, *Regulation of acetic acid production by homo- and heterofermentative lactobacilli in whole-wheat sourdoughs*. Z Lebensm Unters Forsch, 1994. **199**: p. 186-190.
 145. Fonseca BC, B.J., da Silva, TM, dos Passos VF, de Gouvêa PF, Dinamarco TM, Reginatto V, *Butyric acid as sole product from xylose fermentation by a non-solventogenic Clostridium beijerinckii strain under controlled pH and nutritional conditions*. Bioresour Technol Rep, 2020. **10**: p. 100426.
 146. Murty MV, C.T., *Isolation and characterization of xylose- and xylan-utilizing anaerobic bacteria*. Antonie Van Leeuwenhoek, 1989. **55**: p. 153-163.
 147. Gobbetti M, L.P., Minervini F, de Angelis M, Corsetti A, *Arabinose fermentation by Lactobacillus plantarum in sourdough with added pentosans and alpha-L-arabinofuranosidase: A tool to increase the production of acetic acid*. J Appl Microbiol, 2000. **88**: p. 317-324.
 148. Sanz ML, C.G., Gibson GR, Rastall RA, *Selective fermentation of gentiobiose-derived oligosaccharides by human gut bacteria and influence of molecular weight*. FEMS Microbiol Ecol, 2006. **56**: p. 383-388.
 149. Wang H, L.X., Wang Y, Tao Y, Lu S, Zhu X, Li D, *Improvement of n-caproic acid production with Ruminococcaceae bacterium CPB6: Selection of electron acceptors and carbon sources and optimization of the culture medium*. Microb Cell Fact, 2018. **17**: p. 99.
 150. Choi K, J.B., Kim BC, Oh MK, Um Y, Sang BI, *In situ biphasic extractive fermentation for hexanoic acid production from sucrose by megasphaera elsdenii NCIMB 702410*. Appl Biochem Biotechnol, 2013. **171**: p. 1094-1107.
 151. Quesada-Chanto A, A.A., Wagner F, *Microbial production of propionic acid and vitamin B12 using molasses or sugar*. Microbiol Biotechnol, 1994. **41**: p. 378-383.
 152. Dwidar M, K.S., Jeon BS, Um Y, Mitchell RJ, Sang BI, *Co-culturing a novel Bacillus strain with Clostridium tyrobutyricum ATCC 25755 to produce butyric acid from sucrose*. Biotechnol Biofuels, 2013. **6**: p. 35.
 153. Gobbetti M, C.A., Rossi J, *The sourdough microflora. Interactions between lactic acid bacteria and yeasts: Metabolism of carbohydrates*. World J Microbiol Biotechnol, 1994. **10**: p. 275-279.
 154. Babuchowski A, H.E., Glatz BA, *Survey of Propionibacteria for Ability to Produce Propionic and Acetic Acids*. J Food Prot, 1993. **56**: p. 493-496.
 155. Shinohara M, K.M., Tochio T, Kimura S, Koga Y, *Activation of butyrate-producing bacteria as well as bifidobacteria in the cat intestinal microbiota by the*

- administration of 1-kestose, the smallest component of fructo-oligosaccharide. J Vet Med Sci, 2020. **82**: p. 866-874.
156. Cardelle-Cobas A, F.M., Salazar N, Martínez-Villaluenga C, Villamiel M, Ruas-Madiedo P, de los Reyes-Gavilán CG, *Bifidogenic effect and stimulation of short chain fatty acid production in human faecal slurry cultures by oligosaccharides derived from lactose and lactulose*. J Dairy Res, 2009. **76**: p. 317-325.
 157. Taniguchi M, N.H., Takeda O, Kaneko T, Hoshino K, Tanaka T, *Production of a mixture of antimicrobial organic acids from lactose by co-culture of Bifidobacterium longum and Propionibacterium freudenreichii*. Biosci Biotechnol Biochem, 1998. **62**: p. 1522-1527.
 158. Torino MI, T.M., Font de Valdez G, *Citrate catabolism and production of acetate and succinate by Lactobacillus helveticus ATCC 15807*. Appl Microbiol Biotechnol, 2005. **69**: p. 79-85.
 159. Bodie EA, G.N., Schwartz RD, *Production of propionic acid by mixed cultures of Propionibacterium shermanii and Lactobacillus casei in autoclave-sterilized whey*. J Ind Microbiol, 1987. **1**: p. 349–353.
 160. Goswami V, S.A., *Fed-batch propionic acid production by Propionibacterium acidipropionici*. Biochem Eng J, 2000. **4**: p. 121-128.
 161. Jin Z, Y.S., *Extractive fermentation for enhanced propionic acid production from lactose by Propionibacterium acidipropionici*. Biotechnol Prog, 1998. **14**: p. 457-465.
 162. Hsu ST, Y.S., *Propionic acid fermentation of lactose by Propionibacterium acidipropionici: Effects of pH*. Biotechnol Bioeng, 1991. **38**: p. 571-578.
 163. Nisbet DJ, C.D., DeLoach JR, *Effect of mixed cecal microflora maintained in continuous culture and of dietary lactose on Salmonella typhimurium colonization in broiler chicks*. Avian Dis, 1993. **37**: p. 528-535.
 164. Zagrodnik R, D.A., Łężyk M, Oleskowicz-Popiel P, *Enrichment Versus Bioaugmentation-Microbiological Production of Caproate from Mixed Carbon Sources by Mixed Bacterial Culture and Clostridium kluyveri*. Environ Sci Technol, 2020. **54**: p. 5864-5873.
 165. Kenealy WR, C.Y., Weimer PJ, *Production of caproic acid by cocultures of ruminal cellulolytic bacteria and Clostridium kluyveri grown on cellulose and ethanol*. Appl Microbiol Biotechnol, 1995. **44**: p. 507-513.
 166. Cibis KG, G.A., König H, *Isolation of acetic, propionic and butyric acid-forming bacteria from biogas plants*. J Biotechnol, 2016. **220**: p. 51-63.
 167. Jeon BS, K.B., Um Y, Sang BI, *Production of hexanoic acid from d-galactitol by a newly isolated Clostridium sp. BS-1*. Appl Microbiol Biotechnol, 2010. **88**: p. 1161-1167.
 168. Kim BC, S.J.B., Kim S, Kim H, Um Y, Sang BI, *Caproiciproducens galactitolivorans gen. Nov., sp. nov., a bacterium capable of producing caproic acid from galactitol, isolated from a wastewater treatment plant*. Int J Syst Evol Microbiol, 2015. **65**: p. 4902-4908.
 169. Jeon BS, M.C., Kim BC, Kim H, Um Y, Sang BI, *In situ extractive fermentation for the production of hexanoic acid from galactitol by Clostridium sp. BS-1*. Enzyme Microb Technol, 2013. **53**: p. 143-151.
 170. Xu Y, C.Y., Xiang S, Ye K, Bao X, Zhu X, Ge Y, Shi L, Lin M, *Effect of xylitol on gut microbiota in an in vitro colonic simulation*. Turkish J Biochem, 2019. **44**: p. 646-653.
 171. Duarte JC, V.G., Moran PJ, Rodrigues JA, *Microbial production of Propionic and Succinic acid from Sorbitol using Propionibacterium acidipropionici*. AMB Express, 2015. **5**: p. 13.

172. Y, M., *The effect of dietary mannitol on the caecal microflora and short-chain fatty acids in rats*. Lett Appl Microbiol, 1994. **18**: p. 27-29.
173. Wang P, J.Y., Liu S, *Novel fermentation process strengthening strategy for production of propionic acid and vitamin B12 by Propionibacterium freudenreichii*. J Ind Microbiol Biotechnol, 2014. **41**: p. 1811-1815.
174. Ruhel R, C.B., *Use of an osmotically sensitive mutant of Propionibacterium freudenreichii subsp. shermanii for the simultaneous productions of organic acids and trehalose from biodiesel waste based crude glycerol*. Bioresour Technol, 2012. **109**: p. 131-139.
175. Chen F, F.X., Liang JF, Xu H, Ouyang PK, *An oxidoreduction potential shift control strategy for high purity propionic acid production by Propionibacterium freudenreichii CCTCC M207015 with glycerol as sole carbon source*. Bioprocess Biosyst Eng, 2013. **36**: p. 1165-1176.
176. Barbirato F, C.D., Bories A, *Propionic acid fermentation from glycerol: Comparison with conventional substrates*. Appl Microbiol Biotechnol, 1997. **47**: p. 441-446.
177. Coral J, K.S., Porto de Souza Vandenberghe L, Parada JL, Pandey A, Socol CR, *Batch Fermentation Model of Propionic Acid Production by Propionibacterium acidipropionici in Different Carbon Sources*. Appl Biochem Biotechnol, 2008. **151**: p. 333-341.
178. Dishisha T, A.M., Hatti-Kaul R, *Batch- and continuous propionic acid production from glycerol using free and immobilized cells of Propionibacterium acidipropionici*. Bioresour Technol, 2012. **118**: p. 553-562.
179. Dishisha T, I.M., Cavero VH, Alvarez MT, Hatti-Kaul R, *Improved propionic acid production from glycerol: combining cyclic batch- and sequential batch fermentations with optimal nutrient composition*. Bioresour Technol, 2015. **176**: p. 80-87.
180. Dishisha T, S.Å., Lundmark S, Hatti-Kaul R, *An economical biorefinery process for propionic acid production from glycerol and potato juice using high cell density fermentation*. Bioresour Technol, 2013. **135**: p. 504-512.
181. Zhu Y, L.J., Tan M, Liu L, Jiang L, Sun J, Lee P, Du G, Chen J, *Optimization and scale-up of propionic acid production by propionic acid-tolerant Propionibacterium acidipropionici with glycerol as the carbon source*. Bioresour Technol, 2010. **101**: p. 8902-8906.
182. Boyaval P, C.C., Madec MN, *Propionic acid production in a membrane bioreactor*. Enzyme Microb Technol, 1994. **16**: p. 883-886.
183. Wu W, F.Z., *A general inhibition kinetics model for ethanol production using a novel carbon source: Sodium gluconate*. Bioprocess Biosyst Eng, 2013. **36**: p. 1631-1640.
184. Dieterich W, S.M., Zopf Y, *Microbiota in the Gastrointestinal Tract*. Med Sci (Basel), 2018. **6**: p. 116.
185. Krishnan K, C.T., Paster BJ, *A practical guide to the oral microbiome and its relation to health and disease*. Oral Dis, 2017. **23**: p. 276-286.
186. Wood DE, L.J., Langmead B, *Improved metagenomic analysis with Kraken 2*. Genome Biol., 2019. **20**: p. 257.
187. Méric G, W.R., Watts SC, Holt KE, Inouye M, *Correcting index databases improves metagenomic studies*. bioRxiv, 2019: p. 712166.

188. Li D, L.R., Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW, *MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices*. *Methods*, 2016. **102**: p. 3-11.
189. Kang DD, L.F., Kirton E, Thomas A, Egan R, An H, Wang Z, *MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies*. *PeerJ*, 2019. **7**: p. e7359.
190. Nayfach S, S.Z., Seshadri R, Pollard KS, Kyrpides NC, *New insights from uncultivated genomes of the global human gut microbiome*. *Nature*, 2019. **568**: p. 505-510.
191. Chaumeil PA, M.A., Hugenholtz P, Parks DH, *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database*. *Bioinformatics*, 2019. **36**: p. 1925-7.
192. Bowers RM, K.N., Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK et al, *Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea*. *Nat Biotechnol*, 2017. **35**: p. 725-731.
193. Letunic I, B.P., *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation*. *Nucleic Acids Res*, 2021. **49**: p. 293-296.
194. Pascal Andreu V, R.-T.J., Dodd D, Fischbach MA, Medema MH, *The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota*. *Nucleic Acids Res*, 2021. **49**: p. 263-270.
195. Bankevich A, N.S., Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al, *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing*. *J Comput Biol*, 2012. **19**: p. 455-477.
196. T, S., *Prokka: rapid prokaryotic genome annotation*. *Bioinformatics*, 2014. **30**: p. 2068-9.
197. Parks DH, I.M., Skennerton CT, Hugenholtz P, Tyson GW, *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes*. *Genome Res*, 2015. **25**: p. 1043-55.
198. Jolley KA, B.J., Maiden MCJ, *Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications*. *Wellcome Open Res*, 2018. **3**: p. 124.
199. Parks DH, C.M., Waite DW, Rinke C, Skarshewski A, Chaumeil PA et al, *A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life*. *Nat Biotechnol*, 2018. **36**: p. 996-1004.
200. Jain C, R.-R.L., Phillippy AM, Konstantinidis KT, Aluru S, *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries*. *Nat Commun*, 2018. **9**: p. 5114.
201. Segata N, B.D., Morgan XC, Huttenhower C, *PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes*. *Nat Commun*, 2013. **4**: p. 2304.
202. Chun J, O.A., Ventosa A, Christensen H, Arahal DR, da Costa MS et al, *Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes*. *Int J Syst Evol Microbiol*, 2018. **68**: p. 461-466.
203. Goris J, K.K., Klappenbach JA, Coenye T, Vandamme P, Tiedje JM, *DNA-DNA hybridization values and their relationship to whole-genome sequence similarities*. *Int J Syst Evol Microbiol* 2007. **57**: p. 81-91.
204. Richter M, R.-M.R., *Shifting the genomic gold standard for the prokaryotic species definition*. *Proc Natl Acad Sci U S A*, 2009. **106**: p. 19126-19131.

205. Guan X, L.W., Meng H, *A double-edged sword: Role of butyrate in the oral cavity and the gut*. Mol Oral Microbiol., 2021. **36**: p. 121-131.
206. Reitano E, d.A.N., Gavrilidis P, Gaiani F, Memeo R, Inchingolo R et al, *Oral Bacterial Microbiota in Digestive Cancer Patients: A Systematic Review*. Microorganisms, 2021. **9**: p. 2585.
207. Krespi YP, S.M., Kacker A, *The relationship between oral malodor and volatile sulfur compound-producing bacteria*. Otolaryngol Head Neck Surg, 2006. **135**: p. 671-676.
208. Snel J, B.M., Smit B, Noordman W, Tangerman A, Winkel EG, Kleerebezem M, *Volatile sulphur compounds in morning breath of human volunteers*. Arch Oral Biol, 2011. **56**: p. 29-34.
209. Porter SR, S.C., *Oral malodour (halitosis)*. BMJ, 2006. **333**: p. 632-635.
210. van den Velde, S., et al., *Halitosis associated volatiles in breath of healthy subjects*. J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **853**(1-2): p. 54-61.
211. Scully C, G.J., *Halitology (breath odour: aetiopathogenesis and management)*. Oral Dis, 2012. **18**: p. 333-345.
212. Tangerman A, W.E., *Extra-oral halitosis: an overview*. J Breath Res, 2010. **4**: p. 017003.
213. Thorn RM, G.J., *Microbial volatile compounds in health and disease conditions*. J Breath Res, 2012. **6**: p. 024001.
214. Markar, S.R., et al., *Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer*. JAMA Oncol, 2018. **4**(7): p. 970-976.
215. Smith D, S.P., *Studies of interstellar ion reactions using the SIFT technique: Isotope Fractionation*. Acc. Chem. Res., 1992. **24**: p. 414-420.
216. Spanel P and S. D, *Selected Ion Flow Tube - Mass Spectrometry: Detection and real-time monitoring of flavours released by food products*. Rapid Communications in Mass Spectrometry, 1999. **13**: p. 585-596.
217. Hryniuk, A. and B.M. Ross, *Detection of acetone and isoprene in human breath using a combination of thermal desorption and selected ion flow tube mass spectrometry*. International Journal of Mass Spectrometry, 2009. **285**(1-2): p. 26-30.
218. Humphrey SP, W.R., *A review of saliva: Normal composition, flow, and function*. J Prosthet Dent 2001. **85**: p. 162-169.
219. Inui T, P.R., Shah N, Li W, Cisar JO, Wu CD, *Effect of mechanically stimulated saliva on initial human dental biofilm formation*. Sci Rep, 2019. **9**: p. 11805.
220. Iorgulescu, G., *Saliva between normal and pathological. Important factors in determining systemic and oral health*. J Med Life, 2009. **2**: p. 303-307.
221. Bhattarai KR, K.H., Chae HJ, *Compliance with Saliva Collection Protocol in Healthy Volunteers: Strategies for Managing Risk and Errors*. Int J Med Sci, 2018. **15**: p. 823-831.
222. Al-Kateb H, d.L.C.B., Ratcliffe N, *An investigation of volatile organic compounds from the saliva of healthy individuals using headspace-trap/GC-MS*. J Breath Res, 2013. **7**: p. 036004.
223. Monedeiro F, M.M., Ratiu IA, Zmysłowski H, Ligor T, Buszewski B, *VOC Profiles of Saliva in Assessment of Halitosis and Submandibular Abscesses Using HS-SPME-GC/MS Technique*. Molecules, 2019. **24**: p. 2977.
224. Bel'skaya LV, S.E., Shalygin SP, Postnova TV, Kosenok VK, *Identification of salivary volatile organic compounds as potential markers of stomach and colorectal cancer: A pilot study*. J Oral Biosci, 2020. **62**: p. 212-221.

225. Brock B, K.S., Silz J, Trefz P, Schubert JK, Miekisch W, *Monitoring of breath VOCs and electrical impedance tomography under pulmonary recruitment in mechanically ventilated patients*. J Breath Res, 2017. **11**: p. 016005.
226. Fowler SJ, B.-S.M., Xu Y, Goodacre R, Dark PM, *Surveillance for lower airway pathogens in mechanically ventilated patients by metabolomic analysis of exhaled breath: a case-control study*. Thorax, 2015. **70**: p. 320-325.
227. Schnabel R, F.R., Smolinska A, Dallinga J, Boumans ML, Stobberingh E, Boots A, Roekaerts P, Bergmans D, van Schooten FJ, *Analysis of volatile organic compounds in exhaled breath to diagnose ventilator-associated pneumonia*. Sci Rep, 2015. **5**: p. 17179.
228. Bauerfeind P, G.R., Dunn BE, Mobley HL, *Synthesis and activity of Helicobacter pylori urease and catalase at low pH*. Gut, 1997. **40**: p. 25-30.
229. UC, G., *How to interpret hydrogen breath tests*. J Neurogastroenterol Motil, 2011. **17**: p. 312-317.
230. Belluomo I, B.P., Myridakis A, Vadhvana B, Markar SR, Spanel P, Hanna GB, *Selected ion flow tube mass spectrometry for targeted analysis of volatile organic compounds in human breath*. Nat Protoc, 2021.
231. Jers C, K.A., Garg A, Mijakovic I. , *Production of 3-Hydroxypropanoic Acid From Glycerol by Metabolically Engineered Bacteria*. . Front Bioeng Biotechnol., 2019. **7**: p. 124.
232. Yang M, Y.J., Zhang H, Magocha TA, Zabed H, Xue Y, Fokum E, Sun W, Qi X., *Genetically Engineered Strains: Application and Advances for 1,3-Propanediol Production from Glycerol*. . Food Technol Biotechnol., 2018. **56**: p. 3-15.
233. Y., D., *Glycerol metabolism and its regulation in lactic acid bacteria*. . Appl Microbiol Biotechnol. , 2019. **103**: p. 5079-5093.
234. Saito Y, S.T., Nomoto K, Tsuji H. , *Identification of phenol- and p-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites*. FEMS Microbiol Ecol., 2018. **94**.
235. Parada Venegas D, D.I.F.M., Landskron G, González MJ, Quera R, Dijkstra G, Harmsen HJM, Faber KN, Hermoso MA. , *Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases*. . Front Immunol., 2019. **10**: p. 277.
236. Sun Y, O.R.M., *Regulation of bacterial pathogenesis by intestinal short-chain Fatty acids*. . Adv Appl Microbiol., 2013. **85**: p. 93-118. .
237. Dalile B, V.O.L., Vervliet B, Verbeke K. , *The role of short-chain fatty acids in microbiota-gut-brain communication*. . Nat Rev Gastroenterol Hepatol. , 2019. **16**: p. 461-478.

13 APPENDIX

13.1 Search strategy

Total results: 5859 – after duplicates: 4745

Database	Search	Query	Results
Medline	#1	Gastro-oesophag*.mp	5566
	#2	Gastroesophag*.mp	37722
	#3	Gastrooesophag*.mp	420
	#4	Gastro-esophag*.mp	2294
	#5	Esophago-gastr*.mp	1058
	#6	Oesophago-gastr*.mp	609
	#7	Esophagogastr*.mp	16193
	#8	Oesophagogastr*.mp	1435
	#9	Oesophag*.mp	38674
	#10	Esophag*.mp	191769
	#11	Gastr*.mp	795018
	#12	exp Stomach Neoplasms/ or exp Esophageal Neoplasms/	141933
	#13	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12	932527
	#14	exp Neoplasms/	3394264
	#15	exp Adenocarcinoma/	385335
	#16	exp Carcinoma, Squamous Cell/	134529
	#17	Adenocarcinoma.mp	238654
	#18	Squamous cell carcinoma.mp	93606
	#19	Neoplasm.mp	734614
	#20	Cancer.mp	1804438
	#21	#14 OR #15 OR #16 OR #17 OR #18 OR #19 OR 20	3958251
	#22	#13 AND #21	282544
	#23	exp Microbiota/	39323
	#24	Microbiota.mp	61488
	#25	Microbiome.mp	40864
	#26	#23 OR #24 OR #25	80613
	#27	#22 AND #26	2388
Embase	#1	Gastro-oesophag*.mp	8268
	#2	Gastroesophag*.mp	74550
	#3	Gastrooesophag*.mp	930
	#4	Gastro-esophag*.mp	5110
	#5	Esophago-gastr*.mp	1982
	#6	Oesophago-gastr*.mp	1366
	#7	Esophagogastr*.mp	26822
	#8	Oesophagogastr*.mp	2607
	#9	Oesophag*.mp	65283
	#10	Esophag*.mp	325407
	#11	Gastr*.mp	1274851
	#12	exp stomach cancer/ or exp esophagus cancer/	189800
	#13	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12	1486931
	#14	exp malignant neoplasm/	3801324
	#15	exp esophageal adenocarcinoma/ or exp stomach adenocarcinoma/ or exp adenocarcinoma/	246577
	#16	exp esophageal squamous cell carcinoma/ or exp esophagus carcinoma/ or exp squamous cell carcinoma/	205950
	#17	Adenocarcinoma.mp	295924
	#18	Squamous cell carcinoma.mp	207577
	#19	Neoplasm.mp	691776
	#20	Cancer.mp	3816106
	#21	#14 OR #15 OR #16 OR #17 OR #18 OR #19 OR 20	5037378
	#22	#13 AND #21	460805
	#23	exp bacterial microbiome/ or exp microbiome/ or exp intestinal flora/	79693
	#24	Microbiota.mp	66025
	#25	Microbiome.mp	41132
	#26	#23 OR #24 OR #25	114660
	#27	#22 AND #26	3471

13.2 Newcastle Ottawa Scale for Case-Control studies. A maximum of 9 stars can be allocated per study.

Author	Year	Selection				Comparability		Exposure			Total (out of 9)	Ref
		Is the case definition adequate	Representative ness of cases	Selection of controls	Definition of controls	Age	Sex	Ascertainment of exposure	Same method of ascertainment for cases and controls	Non- response rate		
Sjostedt	1985	*	*	*	*	NR	NR	*	*	*	7	[57]
Dicksved	2009	*	*	-	*	NR	NR	*	*	*	7	[44]
Aviles-Jimenez	2014	*	*	*	*	NR	NR	*	*	*	7	[59]
Eun	2014	*	*	*	*	NR	NR	*	*	*	7	[60]
Chen	2015	*	*	*	*	*	*	*	*	*	9	[84]
Hu	2015	*	*	*	*	*	*	*	*	*	9	[61]
Nasrollahzadeh	2015	*	*	*	*	NR	NR	*	*	*	7	[85]
Elliott	2017	*	*	*	*	*	*	*	*	*	9	[87]
Wang	2016	*	*	*	*	NR	NR	*	*	*	7	[62]
Li	2017	*	*	*	*	NR	NR	*	*	*	7	[64]
Castano-Rodrigue	2017	*	*	*	*	NR	NR	*	*	*		[65]
Peters	2017	*	*	*	*	*	*	*	*	*	7	
Ferreira	2018	*	*	*	*	*	*	*	*	*	9	[47]
Sun	2018	*	*	*	*	NR	NR	*	*	*	9	[66]
Hu	2018	*	*	*	*	NR	NR	*	*	*	7	[38]
Wu	2018	*	*	*	*	NR	NR	*	*	*	7	[48]
Hsieh	2018	*	*	*	*	*	*	*	*	*	9	[67]
Coker	2018	*	*	*	*	NR	NR	*	*	*	7	[68]
Gunathilake	2018	*	*	*	*	NR	NR	*	*	*	7	[46]
Liang	2019	*	*	*	*	*	NR	*	*	*	8	[70]
Kageyama	2019	*	*	*	*	NR	NR	*	*	*	7	[71]
Wang	2019	*	*	*	*	*	*	*	*	*	9	[72]
Dong	2019	*	*	*	*	*	*	*	*	*	9	[88]
Qi	2019	*	*	*	*	NR	NR	*	*	*	7	[74]
Peter	2020	*	*	*	*	*	*	*	*	*	9	[76]
Wang	2020	*	*	*	*	NR	NR	*	*	*	7	[91]
Li	2020	*	*	*	*	NR	NR	*	*	*	7	[77]
Wang	2020	*	*	*	*	*	*	*	*	*	9	[92]
Spiegelhauer	2020	*	*	*	*	*	*	*	*	*	9	[78]
Zhao	2020	*	*	*	*	NR	NR	*	*	*	7	[79]
Zhou	2020	*	*	*	*	*	*	*	*	*	9	[93]
Lopetsu	2020	*	*	*	*	NR	NR	*	*	*	7	[94]
Kawasaki	2020	*	*	*	*	NR	NR	*	*	*	7	[95]
Gantuya	2020	*	*	*	*	*	NR	*	*	*	8	[96]
Wu	2020	*	*	*	*	NR	NR	*	*	*	7	[80]
Xu	2020	*	*	*	*	*	*	*	*	*	9	[81]
Dang	2020	*	*	*	*	*	*	*	*	*	9	[82]
		*	*	*	*	NR	NR	*	*	*	7	[83]

13.3 Newcastle Ottawa Scale for Cohort. A maximum of 9 stars can be allocated per study.

Author	Year	Selection				Comparability		Outcome			Total (out of 9)	Ref
		Representative ness of the exposed cohort	Selection of the non- exposed cohort	Ascertainmen t of exposure	Demonstration that outcome of interest was not present at start of study	Age	Sex	Assessmen t of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow-up of cohorts		
Seo	2014	*	*	*	*	*	*	*	*	*	9	[58]
Yamamura	2016	*	*	*	*	*	*	*	*	*	9	[86]
Yu	2017	*	*	*	*	*	*	*	*	*	9	[63]
Shao	2019	*	*	*	*	NR	NR	*	*	*	7	[69]
Chen	2019	*	*	*	*	*	*	*	*	*	9	[73]
Yamamura	2019	*	*	*	*	*	*	*	*	*	9	[89]
Liu	2019	*	*	*	*	*	*	*	*	*	9	[75]
Xu	2020	*	*	*	*	NR	NR	*	*	*	7	[90]
Yuda	2020	*	*	*	*	*	*	*	*	*	9	[55]

13.4 Concentrations (ppbv) of target short chain fatty acids and aldehydes measured with the SIFT-MS in ten participants (S1-S10). Samples: 1 = week one, sample 1, 2 = week one, sample 2, 3 = week one, sample 3, 4 = week two, sample 4, 5 = week two, sample 5. Coefficient of variation percentage, CV%.

Compound	Sample concentration (ppbv)																													
	S1						S2						S3						S4						S5					
	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%
Acetone	320	208	372	216	209	29	88	87	117	99	55	25	132	315	130	145	431	59	136	296	260	198	322	31	286	134	146	405	189	49
Short chain fatty acids																														
Acetic acid	19.0	15.0	11.4	10.3	7.9	34	16.4	5.9	7.3	6.4	6.0	54	7.9	7.4	5.3	4.0	4.9	29	5.1	4.8	4.0	5.3	6.1	15	6.6	4.3	4.2	4.4	3.5	25
Butanoic acid	320	208	372	216	209	29	88	87	117	99	55	25	132	315	130	145	431	59	136	296	260	198	322	31	286	134	146	405	189	49
Hexanoic acid	7.1	9.3	4.4	5.4	3.9	36	10.4	4.9	5.1	3.9	3.0	53	5.1	3.5	3.3	3.2	1.7	36	6.5	4.0	10.5	8.5	8.8	33	4.3	4.2	3.2	5.9	3.8	23
Pentanoic acid	1.3	1.3	0.9	1.4	1.2	17	3.3	1.4	1.6	1.1	1.3	50	1.8	1.4	1.2	1.3	0.8	29	2.6	1.1	1.8	1.2	1.5	37	2.4	1.3	0.9	1.4	1.1	41
Propanoic acid	5.8	4.8	4.7	4.1	3.9	16	8.1	3.3	6.3	4.9	4.1	36	4.6	3.1	2.9	2.6	2.5	27	4.9	3.5	4.2	3.7	3.9	14	7.1	8.2	5.1	6.5	4.2	26
Aldehydes																														
Acetaldehyde	135	172	333	185	201	37	198	178	215	121	219	21	95	49	88	149	43	50	150	297	199	99	673	81	83	51	85	99	116	28
Butanal	329	190	437	214	240	36	104	70	94	111	64	23	121	355	131	146	457	64	180	248	281	198	300	21	231	131	174	372	180	43
Decanal	2.2	2.1	2.6	2.8	2.6	11	19.1	2.9	4.7	2.5	2.3	114	3.2	3.1	3.5	2.5	2.7	14	7.5	3.3	6.2	5.7	3.5	34	6.6	3.4	3.3	3.6	2.9	38
Heptanal	1.2	1.3	1.3	1.1	1.4	9	1.5	1.1	1.0	1.5	1.3	17	1.7	1.7	2.4	1.7	1.6	17	1.5	1.6	2.0	3.5	2.1	38	2.0	1.4	2.0	1.9	1.4	17
Hexanal	1.8	2.7	2.3	2.2	1.7	18	3.4	1.8	2.2	2.3	1.3	35	1.8	1.7	2.9	1.6	1.4	32	2.9	2.3	2.5	2.9	2.0	16	7.8	3.8	3.5	4.9	2.6	44
Nonanal	3.5	4.6	3.6	2.9	3.6	18	7.0	3.3	5.3	3.8	5.2	30	4.3	4.1	4.3	4.0	3.6	8	7.9	7.0	17.5	7.3	6.0	52	4.7	3.1	3.8	4.0	3.6	16
Octanal	2.5	2.6	3.1	3.2	2.8	10	3.6	2.9	3.4	2.7	2.9	12	3.2	2.9	3.4	3.0	2.7	9	4.3	3.0	4.1	4.5	3.0	19	4.9	3.1	4.0	3.9	3.5	18
Pentanal	1.8	2.3	1.7	2.2	1.8	14	4.4	1.5	2.9	1.6	1.6	53	1.7	2.1	1.7	1.7	1.3	18	3.4	2.4	3.5	2.9	3.5	16	5.6	3.6	4.7	7.7	4.1	32
Pronanal	7.1	12.7	3.8	10.4	3.5	54	35.7	13.4	18.4	9.0	12.6	59	17.1	13.2	11.4	9.7	7.8	30	30.3	14.6	31.9	23.4	19.0	31	32.9	12.2	12.6	7.6	15.4	61

Compound	Sample concentration (ppbv)																													
	S6						S7						S8						S9						S10					
	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%	1	2	3	4	5	CV%
Acetone	398	243	401	390	299	21	157	129	264	193	228	28	146	65	83	66	132	39	201	156	222	229	264	19	277	127	140	187	473	59
Short chain fatty acids																														
Acetic acid	8.4	7.5	8.4	8.4	8.0	5	10.4	9.5	8.0	8.2	11.1	14	8.8	6.8	6.1	7.3	7.3	14	7.5	6.6	6.9	8.5	5.7	15	7.1	5.5	4.5	6.6	10.4	33
Butanoic acid	398	243	401	390	299	21	157	129	264	193	228	28	146	65	83	66	132	39	201	156	222	229	264	19	277	127	140	187	473	59
Hexanoic acid	4.0	4.3	3.9	2.9	2.7	20	3.1	3.0	2.8	3.2	2.9	5	3.7	3.5	2.9	2.7	4.9	24	2.3	6.5	3.9	5.8	5.8	35	3.9	3.4	3.2	3.6	5.6	24
Pentanoic acid	1.6	1.7	1.9	2.0	1.2	17	0.9	0.6	0.8	0.7	0.8	15	1.0	1.2	0.7	0.7	0.8	24	0.9	1.2	1.2	1.6	0.9	22	1.1	1.6	0.8	0.9	0.9	32
Propanoic acid	4.1	3.1	3.1	3.4	3.7	12	3.5	4.0	3.3	2.9	3.4	11	3.2	2.8	2.5	2.2	3.4	17	3.1	4.1	4.7	6.4	3.8	28	3.9	4.0	2.9	2.7	8.2	51
Aldehydes																														
Acetaldehyde	63	25	81	47	53	39	4.3	4.4	3.6	4.1	3.3	12	3.3	2.9	3.7	2.5	2.7	16	277	527	393	173	538	41	468	164	250	363	548	44
Butanal	556	240	549	382	277	37	1.7	1.5	1.3	2.0	1.3	19	1.6	1.2	1.5	1.3	1.3	13	232	160	234	225	264	17	216	132	138	263	600	71
Decanal	3.8	3.1	3.1	2.6	2.6	16	1.9	1.7	1.5	2.7	1.3	28	2.9	1.7	2.8	1.6	1.1	40	3.6	6.5	5.5	6.3	5.1	22	3.9	3.1	2.8	3.5	3.4	13
Heptanal	2.8	2.3	3.3	1.9	2.2	21	5.7	7.9	6.0	10.9	7.4	27	6.8	4.7	7.5	4.1	3.8	31	1.7	4.5	5.9	5.4	2.3	48	1.6	2.0	1.6	1.5	1.7	11
Hexanal	3.7	3.0	4.5	2.7	2.0	30	3.1	3.0	2.5	3.1	2.5	11	2.6	2.2	2.8	2.2	1.7	18	2.0	5.5	5.0	7.2	2.3	50	1.9	2.3	1.4	2.2	2.8	25
Nonanal	7.5	5.0	6.1	4.5	2.9	33	2.4	1.7	1.7	2.9	1.1	35	1.1	1.6	2.3	1.4	1.9	28	6.6	13.5	17.9	14.5	10.9	33	11.5	6.3	3.6	6.8	8.7	39
Octanal	5.3	4.2	5.5	4.0	3.4	20	11.3	10.9	10.3	14.0	8.3	19	12.8	13.6	10.8	10.3	12.6	12	2.8	4.3	4.0	4.2	3.4	17	3.2	2.7	2.6	2.1	2.4	15
Pentanal	3.0	1.9	3.6	2.4	2.5	25	47.8	59.7	59.1	35.0	36.6	25	38.2	35.8	61.0	24.8	69.6	41	4.2	3.2	2.9	5.0	3.3	24	2.6	2.8	1.9	2.0	4.4	37
Pronanal	14.9	13.5	14.9	15.7	9.9	17	4.3	4.4	3.6	4.1	3.3	12	3.3	2.9	3.7	2.5	2.7	16	12.1	22.8	17.7	26.9	18.7	28	14.7	15.2	11.2	12.6	16.3	15

13.5 Mixed bacterial profile of gastric cancer associated bacteria across the *in vitro* longitudinal augmentation study

Sample type	Time interval	Taxonomy
Gastric cancer (403)	Baseline	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__Staphylococcus;s__Staphylococcus epidermidis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus vestibularis
	10 minutes	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__Staphylococcus;s__Staphylococcus epidermidis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus vestibularis
	30 minutes	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__Staphylococcus;s__Staphylococcus epidermidis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus vestibularis
	60 minutes	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__Staphylococcus;s__Staphylococcus epidermidis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus vestibularis
	120 minutes	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__Staphylococcus;s__Staphylococcus epidermidis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus vestibularis
Saliva Aerobic (gastric cancer; 403)	Baseline	p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Bacillus amyloliquefaciens p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria elongate p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Aggregatibacter;s__Aggregatibacter segnis p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia mucilaginoso_B p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria flavescens_C p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D sp001679485 p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii
	10 minutes	p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Bacillus amyloliquefaciens p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria elongate p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae

		p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Aggregatibacter;s__Aggregatibacter segnis p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria flavescens_C p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D sp001815355 p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia mucilaginosa_B p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Gemellaceae;g__Gemella;s__Gemella morbillorum p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii
	30 minutes	p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Bacillus amyloliquefaciens p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria elongate p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Aggregatibacter;s__Aggregatibacter segnis p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia mucilaginosa_B p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria flavescens_C p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Gemellaceae;g__Gemella;s__Gemella morbillorum p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria subflava_B p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D sp001679485 p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria perflava
	60 minutes	p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Bacillus amyloliquefaciens p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria elongate p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Aggregatibacter;s__Aggregatibacter segnis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D sp001679485 p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria flavescens_C
	120 minutes	p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Bacillus amyloliquefaciens p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria elongate p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Aggregatibacter;s__Aggregatibacter segnis p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria macacae p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sp000186165 p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Granulicatella;s__Granulicatella sp001071995 p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria subflava_B

		p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D sp001679485 p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia mucilaginosa_B p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Gemellaceae;g__Gemella;s__Gemella morbillorum
Saliva Anaerobic (gastric cancer; 403)	Baseline	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_S p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus sp900104285 p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca_D
	10 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_S p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus mitis

Sample type	Time interval	Taxonomy
Gastric cancer (400)	Baseline	p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A paranthracis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus pneumoniae_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus infantis
	10 minutes	p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A paranthracis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus pneumoniae_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus infantis
	30 minutes	p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A paranthracis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus infantis
	60 minutes	p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A paranthracis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus pneumoniae_D
	120 minutes	p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A paranthracis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus mitis
Saliva Aerobic (gastric cancer; 400)	Baseline	p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseri p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_Ga;s__Neisseria sicca_D
	10 minutes	p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_G p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus sanguinis p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__Rothia dentocariosa

	30 minutes	p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_G
	60 minutes	p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_G p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus sanguinis
	120 minutes	p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Neisseriaceae;g__Neisseria;s__Neisseria sicca_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_G p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus sanguinis
Saliva Anaerobic (gastric cancer; 400)	Baseline	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_W p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis_B
	10 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_W
	30 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_W
	60 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_W
	120 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis_W

Sample type	Time interval	Taxonomy
Gastric cancer (383)	Baseline	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigoritolerans p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lacticaseibacillus;s__Lacticaseibacillus paracasei p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A
	10 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigoritolerans p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lacticaseibacillus;s__Lacticaseibacillus paracasei p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A
	30 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii

		p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigiditolerans p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactocaseibacillus;s__Lactocaseibacillus paracasei p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Actinomycetaceae;g__Pauljensenia;s__Pauljensenia odontolytica_C
	60 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigiditolerans p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactocaseibacillus;s__Lactocaseibacillus paracasei p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Actinomycetaceae;g__Pauljensenia;s__Pauljensenia sp001072465
	120 minutes	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigiditolerans p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactocaseibacillus;s__Lactocaseibacillus paracasei p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella dispar_A p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Actinomycetaceae;g__Pauljensenia;s__Pauljensenia odontolytica_C
Saliva Aerobic (gastric cancer; 383)	Baseline	p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Enterobacter;s__Enterobacter roggenkampii p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Enterobacter;s__Enterobacter himalayensis
	10 minutes	p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Enterobacter;s__Enterobacter roggenkampii
Saliva Anaerobic (gastric cancer; 383)	120 minutes	p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella atypica p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Veillonellaceae;g__Veillonella;s__Veillonella parvula_A p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus gordonii p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus oralis p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Pasteurellaceae;g__Haemophilus_D;s__Haemophilus_D parainfluenzae_K p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Stomatobaculum;s__Stomatobaculum longum p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Actinomycetaceae;g__Pauljensenia;s__Pauljensenia sp000308055 p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Actinomycetaceae;g__Pauljensenia;s__Pauljensenia sp000411415 p__Firmicutes_C;c__Negativicutes;o__Veillonellales;f__Megasphaeraceae;g__Anaeroglobus;s__Anaeroglobus micronuciformis p__Fusobacteriota;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__Fusobacterium periodonticum_D p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__Streptococcus parasanguinis_B p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas;s__Porphyromonas somerae_A p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Atopobiaceae;g__Lancefieldella;s__Lancefieldella sp000564995

13.6 Tyrosine consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.

	Control (n=6)			Oesophageal Cancer (n=6)			Gastric Cancer (n=6)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-tyrosine: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-tyrosine: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-tyrosine Concentration (ppbv)	Fold change	
Short chain fatty acids										
Acetic acid	37.22	34.95	0.91	25.33	13.19	0.62	22.53	24.60	1.29	25-35
Butyric acid	7.08	5.85	0.98	8.60	4.56	0.69	9.77	8.61	0.86	25-35
Hexanoic acid	1.78	1.57	0.86	1.11	0.84	0.65	1.33	0.83	0.92	25-35
Pentanoic acid	4.86	5.44	1.19	3.78	2.48	0.66	3.76	6.10	0.80	25-35
Propanoic acid	22.58	16.44	0.84	17.79	8.49	0.60	31.75	25.60	0.58	25-35
Aldehydes										
Acetaldehyde	40.93	40.44	1.04	22.73	22.3	1.02	29.43	23.93	0.82	25-35
Butanal	5.45	5.16	0.92	4.57	4.03	1.18	5.16	4.22	0.74	25-35
Decanal	2.04	2.11	0.89	1.08	1.27	1.11	1.14	1.08	0.83	25-35
Heptanal	1.46	2.04	1.32	0.70	0.99	1.46	0.95	1.26	1.07	25-35
Hexanal	9.23	11.48	1.12	4.55	6.36	1.25	4.73	5.53	1.26	45
Nonanal	3.99	4.65	1.26	2.34	2.41	0.95	2.38	2.64	1.16	45
Octanal	1.38	1.70	1.11	0.91	1.04	0.94	1.08	0.93	0.79	25-35
Pentanal	2.26	3.21	0.95	1.66	1.26	0.76	2.63	2.91	1.11	25-35
Propanal	17.98	17.12	1.01	8.52	9.05	0.94	11.70	10.51	0.85	25-35
Phenol-alkanes										
1-hydroxy-4-ethylbenzene	0.59	0.58	1.24	0.59	0.48	0.99	0.32	0.61	1.06	35-45
Decane	5.59	8.12	1.22	3.27	3.5	1.16	3.3	4.65	1.27	35-45
Dodecane	1.21	1.21	1.03	0.56	0.63	1.32	0.74	0.93	1.28	35-45
P-cresol	1.24	1.59	0.90	0.92	1.18	1.12	1.72	2.70	1.10	35-45
Phenol	9.01	11.0	1.36	4.65	5.71	1.36	5.41	11.03	1.68	10

13.7 Combined carbohydrates (sugars) consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.

	Control (n=6)			Oesophageal Cancer (n=6)			Gastric Cancer (n=3)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-sugars: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-sugars: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-sugars Concentration (ppbv)	Fold change	
Short chain fatty acids										
Acetic acid	30.5	42.9	1.72	30.1	37.1	2.37	27.9	84.0	4.32	25-30
Butyric acid	6.90	36.0	3.75	9.32	18.1	4.51	9.82	41.0	5.58	5-10
Hexanoic acid	0.66	0.95	1.25	1.38	1.04	0.63	0.91	1.20	1.57	15-20
Pentanoic acid	3.46	3.62	1.12	4.27	2.87	0.73	3.00	1.60	0.55	10
Propanoic acid	13.8	32.8	3.49	13.6	18.5	4.38	14.2	30.3	2.89	20-25
Aldehydes										
Acetaldehyde	23.5	21.6	0.96	25.2	24.1	0.92	14.4	17.5	1.28	10
Butanal	2.45	3.62	0.92	3.86	4.36	1.42	2.59	2.84	1.60	15-25
Decanal	1.15	0.84	0.78	1.22	1.00	0.74	0.66	0.38	0.72	15
Heptanal	1.72	1.46	0.75	1.62	1.34	0.81	1.28	1.14	0.90	15
Hexanal	3.44	4.38	1.36	3.37	4.51	1.60	2.89	3.16	1.11	30
Nonanal	1.73	2.88	1.61	2.90	2.70	1.04	2.48	2.18	0.82	20
Octanal	0.96	1.17	1.02	1.02	0.88	1.02	0.58	0.77	1.28	25
Pentanal	1.46	1.62	1.30	1.97	1.62	0.73	0.85	1.24	2.13	5
Propanal	7.72	9.49	1.37	8.54	10.9	1.75	9.56	12.4	1.35	15-25
Phenol-alkanes										
1-hydroxy-4-ethylbenzene	0.80	0.49	0.89	0.62	0.53	0.72	0.72	0.42	0.66	20
Decane	3.28	5.42	1.54	3.82	4.90	1.28	3.57	4.05	1.04	15
Dodecane	0.41	0.95	1.52	0.79	0.77	0.98	0.62	0.68	1.12	30
P-cresol	0.67	0.87	1.82	0.53	1.18	1.60	0.79	1.12	1.84	10-30
Phenol	2.39	2.96	1.68	2.66	3.21	0.97	4.07	4.47	0.76	15

13.8 Glucose and citric acid consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls and oesophagogastric adenocarcinoma.

	Control (n=6)			Oesophagogastric Cancer (n=3)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-glucose+citric acid: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glucose+citric acid: Concentration (ppbv)	Fold change	
Short chain fatty acids							
Acetic acid	19.0	23.3	1.28	74.3	27.6	0.32	10
Butyric acid	2.88	14.9	5.36	7.61	28.2	6.15	5
Hexanoic acid	1.06	0.62	0.62	1.31	1.30	0.98	5
Pentanoic acid	3.11	2.65	0.99	5.02	3.26	0.65	5
Propanoic acid	8.19	15.5	1.96	46.0	16.54	0.52	10
Aldehydes							
Acetaldehyde	19.7	21.1	0.94	28.0	25.4	0.94	5
Butanal	2.82	2.37	0.83	4.51	4.43	0.78	5
Decanal	0.87	0.96	1.48	1.97	2.35	1.30	20-30
Heptanal	1.80	1.22	0.77	1.85	1.61	0.87	5
Hexanal	4.04	5.94	1.59	3.84	6.22	1.60	30
Nonanal	2.30	2.77	1.23	2.64	3.16	1.21	30
Octanal	1.43	0.96	0.75	1.37	1.43	0.83	30
Pentanal	1.32	1.43	0.89	1.50	1.39	0.91	5
Propanal	6.96	9.41	1.29	10.4	8.49	0.8	15
Phenol-alkanes							
1-hydroxy-4-ethylbenzene	0.52	0.59	1.09	0.90	0.77	0.76	5
Decane	3.92	5.35	1.39	3.28	6.52	1.95	30
Dodecane	0.60	0.81	1.91	0.91	1.00	1.74	25-30
P-cresol	0.87	0.80	0.97	1.11	1.31	1.21	5
Phenol	6.18	7.87	1.56	5.72	7.12	1.38	15

13.9 Glucose oral rinse consumption: Comparison in the concentrations (ppbv) and fold change of volatile organic compounds at the optimal time between three groups; healthy controls, oesophageal adenocarcinoma and gastric adenocarcinoma.

	Control (n=2)			Oesophagogastric Cancer (n=5)			Optimal time (mins)
	Baseline Concentration (ppbv)	Post-glucose: Concentration (ppbv)	Fold change	Baseline Concentration (ppbv)	Post-glucose: Concentration (ppbv)	Fold change	
Short chain fatty acids							
Acetic acid	46.4	118	2.49	38.2	28.3	0.99	10
Butyric acid	8.65	65.4	10.9	8.33	20.1	2.84	5
Hexanoic acid	2.30	1.82	0.94	1.42	1.34	0.94	5
Pentanoic acid	4.45	6.46	1.48	3.73	3.91	0.88	5
Propanoic acid	22.5	169	7.05	32.6	14.5	0.99	5
Aldehydes							
Acetaldehyde	19.6	27.7	1.50	21.4	23.0	0.95	10
Butanal	4.14	3.71	1.37	4.95	6.31	1.22	5
Decanal	1.27	1.09	0.89	1.05	1.44	0.76	5
Heptanal	1.52	1.38	0.90	1.61	1.17	0.80	5
Hexanal	3.91	6.03	1.92	3.62	4.95	1.34	15
Nonanal	3.28	3.00	0.92	3.03	2.60	0.95	5
Octanal	1.53	1.41	0.92	1.11	1.76	1.36	30
Pentanal	1.83	2.26	1.23	2.53	2.05	0.81	10
Propanal	7.58	15.7	2.18	9.06	10.8	0.80	5