Prifysgol Cymru
Aberystwyth
The University of Wales

Clinical Study of Saliva Metabolomics and Microbiomics in Respiratory Diseases

Denisa Asandei (dra3@aber.ac.uk)

Supervisors: Luis A. J. Mur (lum@aber.ac.uk)

Sharon Huws (hnh@aber.ac.uk)

January 2018

IBERS
Athrofa y Gwyddorau Biolegol, Amgylcheddol a Gwledig
Institute of Biological, Environmental and Rural Sciences

Mandatory Layout of Declaration/Statements

Word Count of thesis: DECLARATION	34,616
This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.	
Candidate name	Denisa Ramona Asandei
Signature:	
Date	29.01.2018

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where *__correction services__ have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signature:	
Date	29.01.2018

[*this refers to the extent to which the text has been corrected by others]

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signature:	
Date	29.01.2018

NB: *Candidates on whose behalf a bar on access (hard copy) has been approved by the University should use the following version of Statement 2:*

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by Aberystwyth University.

Signature:	
Date	29.01.2018

ABSTRACT

Human saliva has received substantially less attention as a biofluid for the identification of novel biomarkers for diseases, albeit it reflects most of the compounds found in blood. Similarly, there are few studies on the lung microbiome or metabolome especially assessed through saliva as a biosample. In this clinical research study, it was established that certain metabolites and bacteria in saliva change in particular respiratory diseases, such as chronic obstructive pulmonary disease (COPD) and lung cancer, allowing differentiation from healthy patients.

The metabolomic fingerprints of saliva were derived from Flow Infusion Electrospray Mass Spectrometry. ANOVA of the derived data followed by Principal Component Analysis suggested that fingerprints could differentiate between several respiratory diseases (COPD, lung cancer, asthma, pneumonia, idiopathic pulmonary fibrosis, bronchiectasis, and sarcoidosis) and "healthy" controls. Some individual metabolites were tentatively identified based on accurate mass predictions but more work is required to confirm these.

The microbiomes were assessed by 16S rRNA amplicon sequencing. Analysis at genus level identified 11 genera, which significantly differed between the experimental classes. Assessment of the core microbiome has shown that *Streptococcus* was prevalent in every saliva sample. The COPD Exacerbation class was associated with a loss in a range of bacterial genera as shown by estimates of *a*-diversity. However, these samples displayed an increase in *Staphylococcus*. Other genera were reduced in abundance and this was shown by saliva from patients which were undergoing an exacerbation, had significantly lower bacterial loads when compared to the other categories. This in turn may offer insights with regards to the clinical features of respiratory diseases in relation to the levels of bacterial load during stable conditions or exacerbations.

The metabolomic and microbiomic approaches presented here, both show similar variations with respect to the classification of the pulmonary conditions and controls, which makes human saliva a reliable, cost-effective, non-invasive biofluid with prospective uses in novel methods of screening and diagnosing patients suffering of specific diseases, not only of the pulmonary nature.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to those who have contributed to my MPhil thesis and supported me throughout this past two years.

I am extremely grateful to my main supervisor, Professor Luis Mur, for his incredible kindness and guidance, his patience and valuable knowledge. I would also like to thank Dr Sharon Huws, my second supervisor, for her useful insights and help at various stages of my research.

Besides my advisors, I would like to thank for their assistance and support during the course of my MPhil to Professor Keir Lewis from Prince Philip Hospital, Dr Manfred Beckman, Dr Matthew Hegarty, Dr Ricardo da Costa, Toby Wilkinson, Kieron O'Shea, Adrian Mironas, Thomer Durman, Divya Kattupalli and David Fazakerley from the Institute of Biological, Environmental and Rural Sciences (IBERS). Very special thanks to the research nurses and consultants from Bronglais, Glangwili and Prince Philip Hospitals in the Hywel Dda University Health Board, who have helped me along with patient recruitment and sample collection.

My sincere gratitude is reserved for all patients who agreed to participate in my research project and donate clinical samples even in their most distressed moments, and those who donated samples at Aberystwyth University.

Heartfelt thanks go to my very dear friend and colleague Rachel Paes de Araujo and my friend Eduard Moreno Gil, who have offered me unconditional support and constant encouragements. I would like to thank my parents Dana and Costel Asandei for supporting me throughout my studies all these years.

Finally, I would like to thank Dr Justin Pachebat and Dr Julian Hopkins for taking the time to correct my thesis, their insightful discussion and suggestions about the research.

CONTENTS

LIST OF ABBREVIATIONS

.csv	Comma Separated Values
An	Adenine
AIDS	Acquired Immune Deficiency Syndrome
AIRE	Asthma Insights and Reality in Europe
ANOVA	One-way Analysis of Variance
ARDS	Acute Respiratory Distress Syndrome
	Aspiration
AUC	Area Under the Curve
BALF	Bronchoalveolar Lavage Fluid
BDP	Beclomethasone Dipropionate
BGH	Bronglais General Hospital
BMI	Body Mass Index
bp	base pair
C	Cytosine
CAP	Community-Acquired Pneumonia
CAT	COPD Assessment Test
$CHCl_3$	Chloroform
COPD	Chronic Obstructive Pulmonary Disease
CRP	C-Reactive Protein
CS	Corticosteroids
C_T	Threshold Cycle
CT	Computerized Tomography
DES	DNA Elution Solution
DNA	Deoxyribose Nucleic Acid
E. coli	*Escherichia coli*
EBV	Epstein - Barr virus
ECM	Extracellular Matrix

ETS	Environmental Tobacco Smoke
EUBS-TBNA	Endobronchial Ultrasound Transbronchial Needle
FCTC	Framework Convention on Tobacco Control
FEV_1	Forced Expiratory Volume in 1 second
FIE-MS	Flow Infusion Electrospray Mass Spectrometry
FNIH	Foundation for the National Institutes of Health
FVC	Forced Vital Capacity
G	Guanine
GER	Gastroesophageal Reflux
GGH	Glangwili General Hospital
GINA	Global Initiative for Asthma
GLDC	Enzyme Glycine Decarboxylase
GOLD	Gold Initiative for Chronic Obstructive Lung
HC	Healthy Control
HIV	Human Immunodeficiency Virus
HOMIM	Human Oral Microbe Identification Microarray
HPLC	High Performance Liquid Chromatography
HRQoL	health Related Quality of Life
HsCRP	High-Sensitivity C - Reactive Protein
HTN	Hypertension
IARC	International Agency for Research on Cancer
IBERS	Institute of Biological, Environmental and Rural Sciences
ICU	Intensive Care Unit
IgA	Immunoglobulin A
IHD	Ischemic Heart Disease
IL-1	Interleukin-1
IL-1β	Interleukin-1β
IL-17	Interleukin-17
IL-8	Interleukin-8
IPD	Invasive Pneumococcal Disease
IPF	Idiopathic Pulmonary Fibrosis
IQR	Interquartile Range

ISRCTN	International Standard Randomised Controlled Trials Number
Kb	Kilobase
LDCT	Low-Dose Computerized Tomography
m/z	Mass-to-charge ratio
MeOH	Methanol
MMP	Metalloproteinase
MMP3	Metalloproteinase 3
MMP7	Metalloproteinase 7
MMP8	Metalloproteinase 8
MMP9	Metalloproteinase 9
mMRC	modified Medical Research Council
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
NAC	N-Acetyl Cysteine
NMR	Nuclear Magnetic Resonance
NNK	Nicotine-Derived Nitrosaminoketone
NSCLC	Non-Small-Cell Lung Carcinoma
NTC	Non-Template Control
NTC	Non-Template Control
OPLS-DA	Discriminant Orthogonal Partial Least Squares Regression
OSA	Obstructive Sleep Apnoea
PAB	Prooxidant-Antioxidant Balance
PAH	Polycyclic Aromatic Hydrocarbon
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCT	Procalcitonin
PEF	Peak Expiratory Flow
PLS-DA	Partial Least Squares-Discriminant Analysis
PPH	Prince Philip Hospital
PPS	Protein Precipitation Solution

PPS	Protein Precipitation Solution
PSI	Pneumonia Severity Index
qPCR	quantitative Polymerase Chain Reaction
rRNA	Ribosomal Ribonucleic Acid
RSV	Respiratory Syncytial Virus
rt	retention time
SARS	Severe Acute Respiratory Syndrome
SCLC	Small-Cell Lung Carcinoma
SEWS-M	Salt/Ethanol Wash Solution
SEWS-M	Salt/Ethanol Wash Solution
SHS	Second-Hand Smoke
SIgA	Secretory Immunoglobulin A
T	Thymine
TAE	Tris Acetate Ethylenediaminetetraacetic Acid
TAS	Total Antioxidant Status
TIC	Total Ion Count
TNF	Tumour Necrosis Factor
TNF-α	Tumour Necrosis Factor-α
TNM	Tumour Node Metastases
UK	United Kingdom
V	Volt
VIP	Variable Importance in Projection
WBC	White Blood Cell
WHO	World Health Organization

Clinical Study of Saliva Metabolomics and Microbiomics in Respiratory Diseases

CHAPTER 1 | Causes and Impact of Pulmonary Diseases in the United Kingdom

1.1. | Statistical Overview on Pulmonary Diseases

Pulmonary diseases kill one in five people in the United Kingdom (UK) and cause the NHS great expense, with over £6 billion a year, according to the Burden of Lung Disease Report (2006). Moreover, pulmonary diseases still cause more deaths than ischemic heart disease (IHD) in the UK and lung cancer (LC) continues to be the most prevalent cancer killer, exceeding those linked to breast cancer. On average, death rates from pulmonary diseases in the UK exceed those in the EU. The Burden of Lung Disease report (2006), indicates that respiratory diseases cost the NHS £6.6 billion, out of which £3 billion are expenses for the care system, £1.9 billion for the mortality costs and £1.7 billion for the illness costs.

The major pulmonary diseases in the UK comprise pneumonia, asthma, cystic fibrosis, occupational lung disease (e.g. asbestosis), chronic obstructive pulmonary disease (COPD - an umbrella term covering chronic bronchitis and emphysema), obstructive sleep apnoea (OSA), scarring lung diseases and LC. Surveys of the general population conducted in 2012, on people in the UK living with a lung disease suggest that approximatively 12.7 million people have a history of asthma, COPD or another longstanding respiratory illness (Lung Disease Report, 2006).

One in five men and one in four women consulted a general practitioner (GP) for a respiratory complaint (Burden of Lung Disease, 2006). Additionally, estimates from general practice records show that 8 million people have been diagnosed with asthma, 1.2 million with COPD and more than 150,000 with interstitial lung diseases such as idiopathic pulmonary fibrosis (IPF) or sarcoidosis (Table 1.1).

Table 1.1 | Estimates of people in the UK living with a lung disease
Estimated numbers of people in the UK living with a lung disease based on general practice records (British Lung Foundation, 2012). People that have been diagnosed with a pulmonary disease in the UK over the period of 2008-2012 based on the gender incidence. Asthma diagnoses have indicated the highest incidence in both males and females, followed by COPD with an average of approximately 1,200,000, bronchiectasis and obstructive sleep apnoea (OSA) closely tied, sarcoidosis, lung cancer with 85,796 and idiopathic pulmonary fibrosis (IPF). Table retrieved from the British Lung Foundation official website, 2016.

	Asthma	Bronchiectasis	COPD	IPF	Lung cancer	OSA	Sarcoidosis
Overall	8,028,741	211,598	1,201,685	32,479	85,796	201,411	107,824
Male	3,873,724	88,993	627,019	19,450	45,329	152,074	52,514
Female	4,155,017	122,606	574,666	13,028	40,467	49,337	55,310

British Lung Foundation (2016) have further highlighted how lung diseases are one of the leading causes of death in the UK. Over the course of 2008 – 2012, pulmonary diseases were responsible for 20% of all deaths. In 2012, there were 114,225 deaths caused by lung diseases in comparison with 158,383 deaths from cardiovascular diseases (Table 1.2). Despite the fact that the number of deaths from cardiovascular diseases is greater than lung diseases, over the 5-year period (2008 – 2012) the proportion of deaths caused by cardiovascular diseases has been decreasing. The predominant lung diseases in the UK in terms of the highest numbers of mortalities are lung cancer and COPD (Figure 1.1). Both conditions are very often associated with tobacco smoking, which is also a risk factor for pneumonia, another leading cause of

death. During 2012, lung cancer was responsible for 6.2% of all UK deaths, COPD for 5.3% and pneumonia for 5.1%.

Pneumonia, arises mainly from two highly contagious bacterial strains that register 25.3% deaths and pulmonary fibrosis, the lung scarring condition caused by exposure to environmental pollutants and certain medicines, 4.6%. Occupational lung diseases, e.g. mesothelioma which can arise from exposure to asbestos fibres, have increased to a percentage of 2.1% (British Thoracic Society, 2006; British Lung Foundation, 2016).

Table 1.2 | Prevalence of death causes in the UK during 2008-2012
UK deaths by cause in all age groups during 2008 – 2012 (British Lung Foundation, 2012). The rates of mortality in both men and women during the period of 2008-2012 of the most life-threatening diseases in the UK, indicating cardiovascular diseases with percentages ranging between 28-32%, lung diseases with 20% and other diseases with percentages of 27% or 29%. Table retrieved from the British Lung Foundation, 2016.

	Numbers of deaths and % of all deaths									
	2008		2009		2010		2011		2012	
All causes, all ages	577,398		557,366		559,518		550,029		566,924	
Lung diseases	115,928	20%	111,775	20%	110,961	20%	110,921	20%	114,225	20%
Cardiovascular diseases	184,518	32%	177,282	32%	175,320	31%	156,745	28%	158,383	28%
Non-respiratory cancers	121,319	21%	120,678	22%	123,943	22%	125,521	23%	127,968	23%
Other diseases	155,633	27%	147,631	26%	149,294	27%	156,842	29%	166,348	29%
Respiratory deaths	115,928		111,775		110,961		110,921		114,225	
Lung cancer	35,333	6.1%	35,071	6.3%	34,941	6.2%	35,238	6.4%	35,419	6.2%
COPD	28,344	4.9%	26,843	4.8%	27,164	4.9%	28,084	5.1%	29,776	5.3%
Pneumonia	32,282	5.6%	29,909	5.4%	28,405	5.1%	28,381	5.2%	28,952	5.1%
Idiopathic pulmonary fibrosis	3,964	0.7%	3,977	0.7%	4,323	0.8%	4,992	0.9%	5,292	0.9%
Lung diseases due to external agents	3,728	0.6%	3,466	0.6%	3,751	0.7%	3,756	0.7%	4,171	0.7%
Bronchiectasis	1,150	0.2%	1,214	0.2%	1,218	0.2%	1,332	0.2%	1,567	0.3%
Asthma	1,205	0.2%	1,134	0.2%	1,147	0.2%	1,168	0.2%	1,246	0.2%
Sarcoidosis	137	<0.1%	148	<0.1%	149	<0.1%	159	<0.1%	170	<0.1%
Cystic fibrosis	122	<0.1%	147	<0.1%	107	<0.1%	116	<0.1%	111	<0.1%

There is a strong link between age, gender, regional variation, social deprivation and inequality with lung disease (Burden of Lung Disease, 2006). Approximatively 44% of all respiratory disease deaths can be related to social class inequalities in comparison with 28% from IHD. For instance, men aged 20 – 64 labouring in unskilled manual occupations are predisposed 14 times more to die from COPD, than men employed in professional positions.

The burden of respiratory diseases affects individuals and their families, producing high costs to the health care system as well (Office of Disease Prevention and Health Promotion, 2014). The need of new technologies that promote respiratory health through better prevention, detection, treatment and even education have become a priority in the medical field and a subject undergoing intense study in the research area.

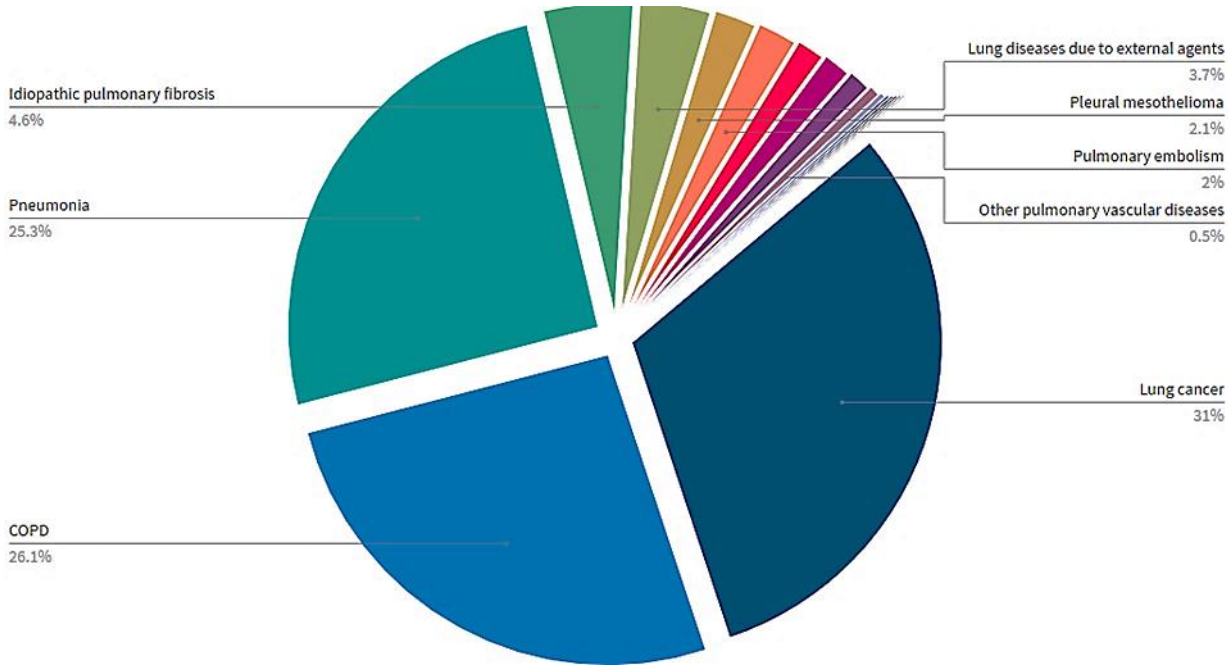


Figure 1.1 | UK deaths from lung diseases (2008-2012)
Lung cancer (LC) primes with 31%, followed by chronic obstructive pulmonary disease (COPD) 26.1%, pneumonia with 25.3%, idiopathic pulmonary fibrosis (IPF) 4.6% and other occupational lung diseases 3.7% or due to environmental pollutants and certain medicines 2.1% (British Lung Foundation, 2012).

1.2 | Chronic Obstructive Pulmonary Disease (COPD)

Chronic obstructive pulmonary disease (COPD) is a progressive lung disease, typically characterised by gradual reduction of the airflow into the lungs caused by the inflammation of the air passages and damage to the lung tissue (Health and Safety Executive, 2016). COPD is a life-threatening heterogeneous disease, an umbrella term that comprises the well-known respiratory conditions, chronic bronchitis and emphysema.

Chronic bronchitis affects the airways (bronchi) and emphysema affects the air sacs (alveoli) in the lungs as shown below in Figure 1.2 (British Lung Foundation, 2012).

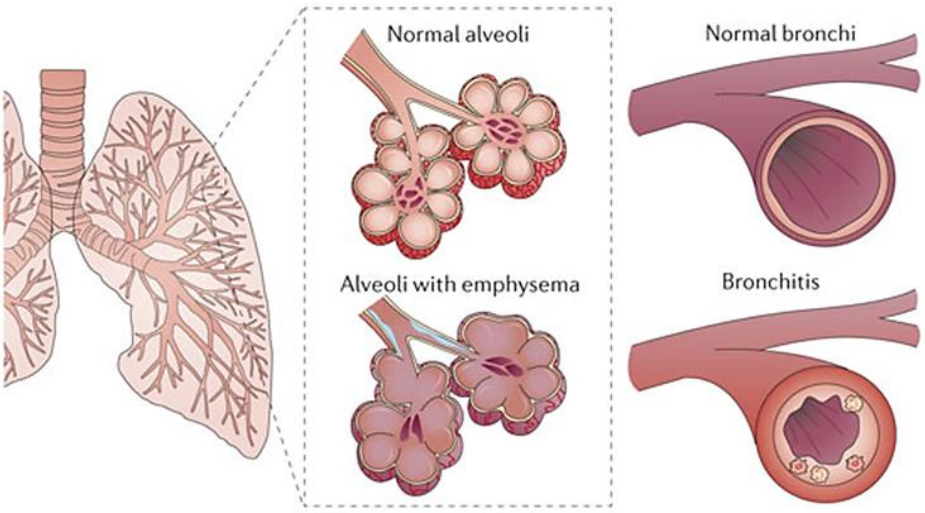


Figure 1.2 | Pathology of the lung affected by emphysema or chronic bronchitis
Emphysema and chronic bronchitis are subdivisions of the respiratory condition COPD, affecting the bronchi and the alveoli in the lungs (Houghton, 2013).

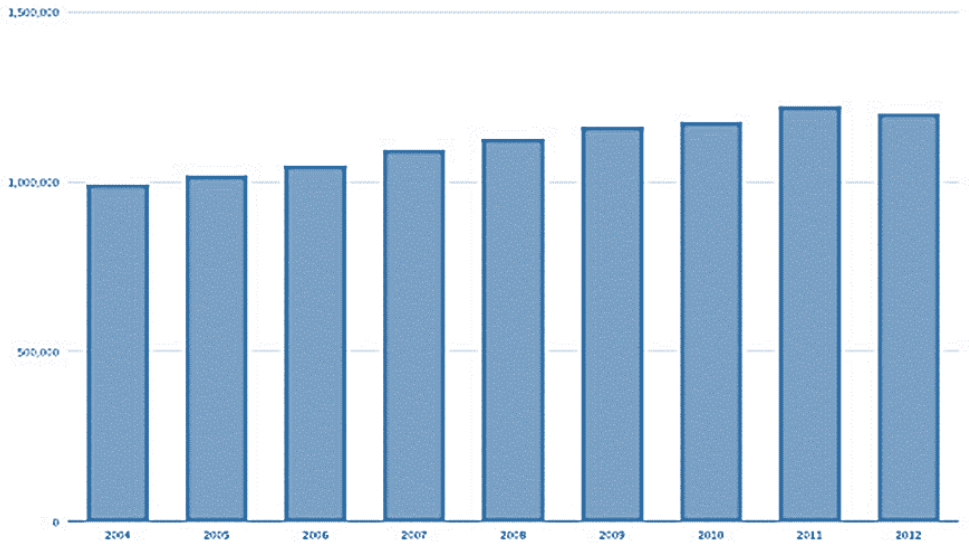
1.2.1 | Statistics in COPD

COPD is common in later life, it becomes apparent after 40 or 50 years of age (Halbert *et al.*, 2006; WHO, 2016). Over the period of 2008-2012, around a million individuals in the UK have been estimated to suffer from COPD (Table 1.3), and over 25,000 deaths are recorded each

year (Health and Safety Executive, 2016). Global estimates (WHO, 2015) show that approximatively 3 million deaths were caused by COPD in 2015 only, which accounts for 5% of all deaths globally. In the UK, the number of individuals diagnosed with COPD has increased by 27% in the last decade (Table 1.3) which potentially means that more undiagnosed cases are being discovered or the condition is becoming more widespread (British Lung Foundation, 2012).

Table 1.3 | Estimated cases of individuals diagnosed with COPD over the period of 2008-2012

Newly diagnosed cases of COPD have increased from under 1600 to approximatively 2000 in the last decade, rising the incidence as well by 9% for the period of 2008-2012 only (British Lung Foundation, 2012).



1.2.2 | Causes of COPD

The vast majority of COPD cases are mainly caused by tobacco smoking, although specific occupational exposures can also represent significant risk factors (Salvi and Barnes, 2010). Moreover, the morbidity of COPD represents a serious social and economic burden and the Global Initiative for Chronic Obstructive Lung Disease (GOLD, 2014) claims that these increase with patient age. It is difficult to assess the precise figures of COPD morbidity, whilst comorbidities interfere with the respiratory condition. Mathers and Loncar (2006), argue that by 2030, COPD will become the fourth leading cause of mortality globally, fact that will employ major consequences worldwide.

1.2.3 | Causes of COPD – Smoking

Since tobacco smoke represents the highest risk factor for COPD, the prevalence in current and ex-smokers is greater compared to non-smokers (Halbert *et al.*, 2006; WHO, 2016). According to the World Health Organization (WHO, 2016), COPD cases are likely to increase in the years to come, caused by the prevailing smoking incidence and the aging population. For this reason, a method to prevent developing COPD in a smoker is either by avoidance or early cessation of tobacco smoking. The WHO Framework Convention on Tobacco Control (WHO-FCTC) (2017) is trying to implement effective strategies of tobacco control and promote non-smoking globally.

Previously, COPD cases were more common in men, compared to women, due to the higher levels of tobacco smoking (Halbert *et al.*, 2006). However, the aforementioned is more relevant among high-income countries. Considering that frequent exposures to household biomass fuel

smoke used for cooking and heating by women in low-income countries are extremely high, COPD now sets both genders at an almost equal level (WHO, 2016).

1.2.4 | Causes of COPD – Occupational Dusts and Fumes

Almost 90% of COPD deaths are recorded in low- and middle-income countries due to indoor air pollution from solid biomass fuel smoke exposure (Salvi and Barnes, 2010; Chakraborty, Mondal and Datta, 2014). Workplace exposures to certain fumes and chemicals including isocyanates, polycyclic aromatic hydrocarbons (PAHs) and welding fumes containing iron with additive metals like titanium, cobalt or copper, and dusts consisting of asbestos, coal, grain and silica (Blanc and Toren, 2007) have been confirmed to contribute to the development of COPD.

A recent evaluation of a Biobank study in the UK on the variation concerning the prevalence of COPD caused by certain occupational exposures, has concluded that former coal miners were predominant and linked to numerous cases of chronic bronchitis and emphysema (Health and Safety executive, 2016). Nowadays, the annual number of cases has decreased as there are less coal miners subjected to long periods of underground exposure.

1.2.5 | Pathogenesis of COPD

COPD is an irreversible disease caused by exposure to certain irritants and develops slowly by triggering an inflammatory response in the respiratory tract. The most common symptoms that COPD present are exacerbations characterised by dyspnoea (breathlessness), chronic cough and production of sputum (phlegm) (WHO, 2016). As the condition progressively worsens, daily activities become difficult, to the point that walking up a short flight of stairs turns into a

strenuous challenge. It is very common during episodes of acute exacerbations for inflammatory events in the respiratory tract to increase. This may be due to colonisation of the lower respiratory tract by bacteria or viruses (Wedzicha and Seemungal, 2007). Moreover, a study conducted by Donaldson and Wedzicha (2014), indicated that seasonal variation represents a factor that can influence the COPD exacerbations as well. The seasonality of exacerbations varies with the geographical latitude, since greater cases are registered in countries with more temperate climates, where people are exposed more frequently to cold, damp conditions. This in turn, increases the level of bacterial load in the lungs of individuals suffering from COPD, as well as the increased prevalence of respiratory viral infections.

Garcha *et al.* (2012) indicated an evident increase in the level of bacterial load in COPD patients with exacerbations compared to COPD patients at mild or moderate stages. This suggests that the microbiome – which describes the microbiological components in a population - plays an important role in a patient's susceptibility to increased airway inflammation. Therefore, characterisation of the COPD microbiome could provide a better understanding on how the immune system influences the condition and could influence the development of better clinical strategies.

Another aspect that plays a significant role in the symptoms of COPD is oxidative stress. Stanojkovic *et al.* (2011) considered the lung functions when assessing the oxidative stress parameters and inflammatory markers in 74 patients suffering from severe COPD exacerbations in some with ischaemic heart disease (IHD) as co-morbidity compared to 41 healthy subjects. Results indicated a low total antioxidant status (TAS) and an elevated prooxidant-antioxidant balance (PAB) in COPD patients compared with individuals from the healthy control group. In addition, COPD patients with positive IHD recorded higher PAB levels in comparison to the negative IHD patients, which confirmed the Alamdari *et al.* (2008) observation that PAB represents a cardiovascular risk factor as well. The Stanojkovic *et al.*,

(2011) study, validated PAB as an oxidative stress marker linked with COPD exacerbations. This offered more insight on how to modulate exacerbations by medical treatment. Subsequently, the inflammatory markers; C-reactive protein (high sensitivity C-reactive protein hsCRP) and the neutrophils percentages, were shown to be higher in COPD patients than in controls. Interestingly, there was a significant correlation between PAB and hsCRP (Pearson's correlation $r = 0.388$, $P < 0.05$) implying a possible link amid oxidative stress and inflammatory markers (Stanojkovic *et al.*, 2011).

Oxidant-antioxidant imbalance and elevated inflammatory markers in COPD-IHD are both common symptoms of end-stage diseases, which belong to the two top groups in causes of mortality as shown in Chapter 1.1, Table 1.2 (Doehner *et al*., 2009; von Haehling *et al*., 2009). This commonality in high oxidant/antioxidant levels and elevated systemic inflammation particularly in patients with both severe COPD exacerbations and IHD (or chronic heart failure) suggests that these diseases are mutually influenced. Indeed, IHD could develop gradually due to the influence of COPD. Therefore, antioxidant therapy constitutes a great form of auxiliary treatment for COPD exacerbations with regards to novel clinical therapeutic interventions.

1.2.6 | Assessment and Diagnosis of COPD

In the case that a person is an active or former smoker or has a history of exposure to other types of COPD risk factors and experiences symptoms such as shortness of breath and chronic cough with sputum production, the next crucial step is to carry a spirometry test (measurement of breath) to confirm the definitive clinical diagnosis.

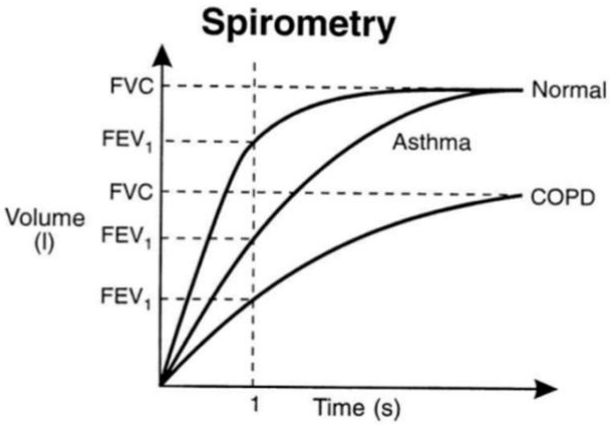


Figure 1.3 | Spirometry test (FEV_1/FVC) standards between normal healthy lungs and COPD- or asthma-affected lungs
COPD diagnosis is usually confirmed following the spirometry test (breathing test) which measures the breath capacity and the speed of forcible exhaled air. A ratio of the FEV_1/FVC (Forced expiratory volume in 1 second/Forced vital capacity) result of less than 0.70 confirms that an individual may be suffering of COPD (Adawy, 2016).

The post bronchodilator test measures, through spirometry, the percentage between the forced expiratory volume in one second (FEV_1) over the forced vital capacity (FVC). Therefore, the FEV_1/FVC ratio, also termed as the FEV_1% of predicted, provides the patient's natural respiratory capacity status (Global Initiative for Chronic Obstructive Pulmonary Disease, 2014). Spirometry test results are recorded before and after bronchodilator administration in order to establish the reversibility of airway obstruction. The bronchodilator reversibility test, along with the past medical history of a patient, is also significant for differentiating the similarities between COPD from asthma, as shown above in Figure 1.3 (Adawy, 2016).

Based on the outcome of FEV_1/FVC ratio, COPD -linked limitation of the airflow can be classified from mild or GOLD I, moderate or GOLD II, severe or GOLD III to very severe or GOLD IV. As it can be observed in Tables 1.4 and 1.5 (new GOLD stages of COPD), the more the airflow limitation worsens the more the FEV_1 value declines.

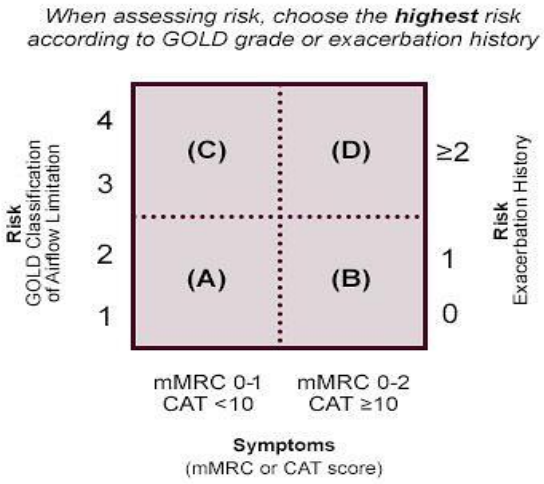
Table 1.4 | Classification of COPD stages according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD)
COPD is classified into four stages depending on the severity of the symptoms that a patient presents. The staging criteria is established by the FEV1 of at 70% or less of FVC ratio (FEV1/FVC) in accordance with the patient's medical history and other symptoms (table retrieved from Respiratory Care and Sleep Medicine, 2016).

Stage I	Mild COPD	$FEV_1/FVC < 0.70$	$FEV_1 \geq 80\%$ normal
Stage II	Moderate COPD	$FEV_1/FVC < 0.70$	FEV_1 50%–79% normal
Stage III	Severe COPD	$FEV_1/FVC < 0.70$	FEV_1 30%–49% normal
Stage IV	Very Severe COPD	$FEV_1/FVC < 0.70$	$FEV_1 < 30\%$ normal, or < 50% normal with chronic respiratory failure*

Although spirometry is the principal technique for diagnosing COPD and staging its severity, Jones, (2009) showed patients with the same FEV_1/FVC ratios, but with different backgrounds varied regarding their quality of life. This suggests that the risks of comorbidities associated with the condition need to be taken into account when screening and diagnosing COPD patients, because the quality of life can have a major influence on the patient's health status.

Furthermore, an evaluation tool that can aid in the screening of the respiratory condition is the COPD Assessment Test (CAT), which measures the influence of COPD on a patient's life based on the changes in health status throughout time (Jones *et al.*, 2009; CAT COPD Assessment Test, 2012). CAT uses a questionnaire composed of eight items that assess the severity of a variety of respiratory symptoms common to COPD, in relation to the patient's health status. At the end of the COPD assessment, every patient obtains a score which is intended to provide insight of the effects that COPD symptoms have on the patient's life. This can be further investigated through consultation and/or intervention. The CAT score should corroborate with the patient's FEV_1, frequency of exacerbations and co-morbidities.

Table 1.5 | New classification system for COPD severity stages according to GOLD guidelines 2011
The new GOLD (Global Initiative for Chronic Obstructive Lung Disease) stages for COPD according to the severity of the airflow limitation, based on Post-bronchodilator FEV_1 results. The GOLD test assessment is based on the CAT or mMRC (modified Medical Research Council) tests. Image retrieved from COPD Management and the 2011 GOLD Guidelines - Peer-reviewed Newsletter, 2013.



Additionally, patients at a more severe stage of COPD have frequent exacerbations and are more inclined to get higher CAT scores, compared to patients at a milder stage of COPD (Jones *et al.*, 2009).

However, crucially, the CAT questionnaire is not suitable for screening, as the diagnosis and severity stage of COPD can only be achieved through spirometry and/or pulmonary function assessment (Hanania *et al.*, 2010).

The new COPD GOLD staging was derived due to the validity of the mMRC test, which measures the dyspnoea- associated disability. The mMRC test correlates with the health status and is able to predict the risk of mortality. Moreover, Nishimura *et al.* (2002) showed that dyspnoea severity correlates more with mortality than the disease severity based on FEV_1. As a result, both CAT and mMRC breathlessness tests are used in clinical settings, and this allows them to be easily integrated in routine clinical practice.

1.2.7 | Treatment and Management of COPD

A key solution to improve the management of COPD consists in adopting a healthy life-style, which can decrease the risk of the disease progression, lessen exacerbations and mortality (GOLD, 2013). Smoking cessation has also proven to be an efficient prevention method as it reduces the inflammatory damage that cigarette smoke is causing to the lungs (GOLD, 2014).

However, there is no known medication that can prevent or at least reverse the progressive lung function deterioration that COPD causes (Cameron, 2015).

Thus, pharmacological and non-pharmacological treatment approaches focus on dealing with COPD symptoms, improve lung function and quality of life. Pharmacological therapy, otherwise known as the cornerstone management for COPD, includes anticholinergics, long-acting bronchodilators (β_2-agonists) and the most used, corticosteroids available in different forms including tablets, injections, inhalers or as lotions and creams (Gold Initiative for Chronic Obstructive Lung Disease, 2014; Respiratory Care and Sleep Medicine, 2016).

It is required for every COPD treatment to be chosen according to every patient's GOLD stage and comorbidities, otherwise it can impose serious complications, i.e. high risk of developing pneumonia after corticosteroid use. Therefore, increased use of rescue medication may aggravate the disease (Donaldson *et al*., 2002).

Non-pharmacological types of medication for COPD include pulmonary rehabilitation, oxygen therapy or pulmonary transplantation (Global Initiative for Chronic Obstructive Lung Disease, 2014).

The greatest challenge in managing COPD in patients at severe GOLD stages is prescribing treatment for the associated frequent exacerbations. This is due to the fact that the source of

exacerbations is always poorly defined and consequently, their impact cannot be minimised through targeted treatment (Respiratory Care and Sleep Medicine, 2016).

Triggers for COPD exacerbations include bacterial or viral changes or infection but also such as temperature changes (Wedzicha and Seemungal, 2007). So far, the most effective treatment for COPD exacerbations has proven to be the antibiotic use. Antibiotic treatment reduced the frequency of exacerbations, with enhanced results especially in patients with severe COPD than in those with moderate COPD.

Comprehending the lung microbiome in COPD patients at an early GOLD stage, may offer new insights regarding the causes that trigger acute exacerbations in patients at more severe stages.

It is important to note that COPD constitutes a driving factor in lung cancer (Durham and Adcock, 2015). There are fundamental predispositions such as increased oxidative stress, repression of DNA repair mechanisms or cellular proliferation, that could determine COPD to progress into lung cancer, or occur as co-morbidities.

1.3 Lung Cancer

It is now well established that cancer arises from active changes in the genome, causing instability and mutations that activate oncogenes and tumour-promoting inflammation leading to limitless replicative potential (Hanahan and Weinberg, 2000; Bishop and Weinberg, 1996). However, cancer in itself can be defined more as a cluster of diseases, which share similar characteristics, such as unregulated cell growth that eventually leads to the metastasis from the primary site of the tumour to other sites in the body (Cameron, 2015). Hanahan and Weinberg (2000) have defined these into a few indicative hallmarks, which are shared by most, and likely, all cancers (Figure 1.4). These hallmarks outline the fact that multistep processes that manifest

genetic alterations trigger the progressive transformation of “normal-working” human cells into highly malignant derivatives. A better understanding of these hallmarks could provide greater results in all areas of cancer research by improving the cause, detection, epidemiology (incidence, distribution and possible control of disease) and treatment (Hanahan and Weinberg, 2000).

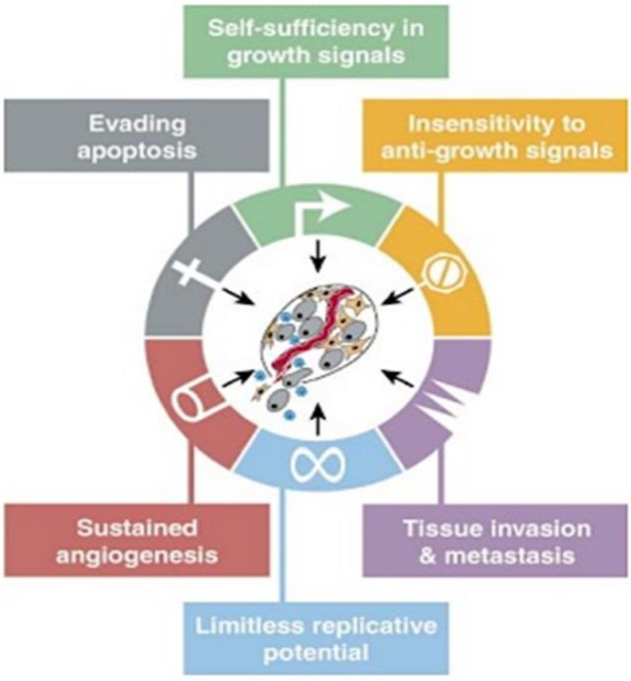


Figure 1.4 | The general hallmarks of cancer throughout the formation of tumorigenesis
Acquired characteristics of most, and likely, all cancers as suggested by Hanahan and Weinberg (2000), that present a general set of functional abilities during the process of their development, although through various automatic strategies. Figure retrieved from Hanahan and Weiner (2000).

1.3.1 | Statistics in Lung Cancer

In the United Kingdom, one of the high-income countries (World Bank Ranking list of economies, 2016), the leading causes of cancer deaths are lung, bowel, breast and prostate cancers which account for 46% of all cancer deaths. Out of almost half (46%) of all cancer

deaths, more than a fifth are due to lung cancer, with 130 new cases diagnosed every day (Cancer Research UK, 2014).

Lung cancer is the leading cause of death in men and second only to breast cancer in women, with an increased incidence rate in people aged 75 and over (Cancer Research UK, 2014).

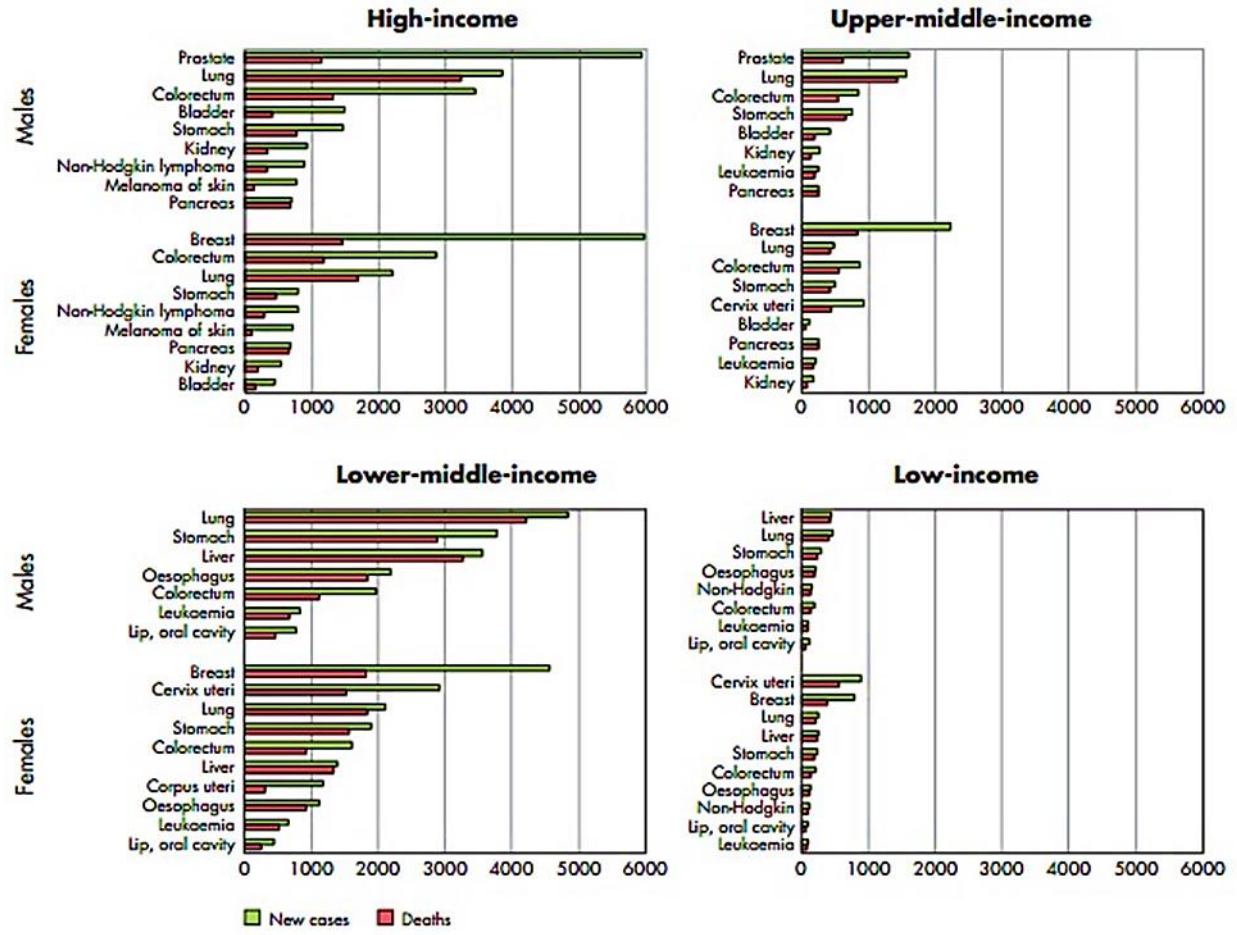
In countries with emerging economies (low-to-middle-income-countries) such as China and India, the incidence (new cases) of lung cancer and deaths produced by the disease are rapidly rising due to the accelerated industrialization and urbanization, in association with incredible changes in life style and environment combined with the aging population (Bode *et al.*, 2015). The risk factors that trigger lung cancer remain the same as in low- and middle-income countries, plus the addition of the eminent four shared behavioural factors that include tobacco use, unhealthy diet, physical inactivity, exposure to radon gas and asbestos, and the harmful use of alcohol (World Health Organization (WHO), 2010).

In lines with such environmental risk factors, studies on migrant communities (Jemal *et al.*, 2010) have revealed that cancer rates in the descendants change towards the cancer rates of the host country. Thus, adopting an "unhealthy western lifestyle", such as smoking and bad dieting to name a few, can eventually overcome the genetic differences between populations (Jemal *et al.*, 2010). Consequently, lung cancer diagnoses have increased dramatically in the last decades and are predicted to become a worldwide major cause of morbidity and mortality in future years. As shown in Table 1.6, despite the geographical variation in cancer distribution, lung cancer is present within the highest numbers of new cases and deaths in both males and females according to the World Bank income groups of countries (WHO, 2010).

Therefore, lung cancer is becoming one of the commonest types of cancer in both sexes, regardless of the financial status or geographical area. Furthermore, whilst the focus falls on lung cancer treatment and genetic susceptibility profiling, a greater interest is arising on developing methods of identifying lung cancer at an earlier stage.

Table 1.6 | Annual number of new cases and deaths for the ten most common types of cancer in accordance with the income groups and gender

High-income countries: USA, Japan, Germany, France, United Kingdom, Italy, etc.; upper-middle-income countries: China, Brazil, Russia, Georgia, Iran, Iraq, etc.; lower-middle-income countries: India, Egypt, Nigeria, etc.; low-income countries: Ethiopia, Senegal, Sierra Leone, Zimbabwe, etc. Lung cancer seems to prevail despite the geographical area or financial status, permanently situated on the first, second or third place in the commonest types of cancer (Tables retrieved from WHO, 2012).



1.3.2 | Causes of Lung Cancer – Smoking

Our present understanding of LC as a disease relies on the prolonged exposure of pulmonary DNA to an array of metabolically activated carcinogens, which are further eliciting changes in key genes regulating cell-growth. Lung cancer statistics suggest that 30% of cases are linked to cigarette smoking (Global Status Report on Noncommunicable Diseases, 2010), typically in the more developed countries. Additionally, smoking is one of the leading causes of lung cancer (International Agency for Research on Cancer IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, 2002), with the foremost indicator of risk being the duration (measured in pack years) rather than quantity of smoked tobacco (Cornfield *et al*., 2009).

In the UK, LC mortalities indicate a high percentage, ranging between 80 – 90%, and are related entirely to smoking (Rossi *et al*., 2005; Hecht, 2002). Currently, 4000 chemicals have been identified in cigarette smoke, with more than 60 recognized as carcinogens by the International Agency for Research on Cancer (Hoffman *et al*., 2001). With respect to LC, the most threatening and harmful carcinogens are PAHs (Polycyclic Aromatic Hydrocarbons) and nicotine-derived nitrosaminoketone (NNK) (Hecht, 1999). Although, nicotine is not recognised as a carcinogen, it can still induce tumours under certain particular conditions (Schuller *et al*., 1995; Hecht *et al*., 2000).

However, the major role of nicotine is through addiction which motivates people continue to smoke despite its known adverse effects on health (Hecht, 2002). Although, each dose for every carcinogen per cigarette is relatively small, the cumulative dose during a smokers' lifetime is substantial. For this reason, cigarettes embody a harmful nicotine device with more than 60 carcinogens in each puff (Hecht, 2000).

Moreover, it has been acknowledged (Wistuba *et al*., 1997; Hecht, 1999; Hanahan and Weinberg, 2000) that, particularly in smokers, metabolically activated carcinogens are directly

associated with progressive mutations in the genome sequence. Wei *et al.* (2000) came to a similar conclusion in a study involving 732 clinical patients, half suffering from LC and half without (controls). This confirmed there were statistically significant differences between the DNA repair capacities of patients suffering from lung cancer in comparison with the control group.

As only a minority of the smoking population develops lung cancer, the probability of being susceptible to the disease could differ from individual to individual. There are various factors, which could potentially induce susceptibility to lung cancer in smokers. These factors range from carcinogen uptake, metabolic stimulation and detoxification, DNA repair proficiency, apoptosis and varying outcomes on genes included in signal transduction pathways and regulation of the cell cycle (Hecht, 1999; Perera, 1997).

Differences in carcinogen metabolic stimulation and detoxification, bring to attention PAH, the most broadly studied carcinogen (Phillips, 1983; Hecht, 1999). Accordingly, PAH, is the prevalent cause of the carcinogenic effect regarding occupational exposure to soots, coal tars and other related materials (Phillips, 1983). Additionally, previous studies in this field have revealed variation amongst individuals concerning the metabolism and DNA-adduct formation by PAHs (Bartsch *et al*., 2000; Smith *et al*., 2001). Therefore, humans differ in respect to the process of PAH activation and detoxification. Moreover, individuals who are exposed to PAH compounds excessively are at greater risk because of its carcinogenic effects, rather than persons who efficiently detoxify those (Smith *et al*., 2001).

In the context of human tobacco-carcinogen uptake, it is necessary to make advances in the comprehension of carcinogen dose and its mechanism of carcinogenesis. This in turn, can provide the biochemical foundation for investigating the link between exposure to tobacco smoke and LC (Hecht, 2002). Following Spiegelhalder and Bartsch's research study (1996),

NNK is classified as a systemic pulmonary carcinogen that generates lung tumours, as it reaches the lung via cigarette smoke. Consequently, NNK carcinogens can generate the development of adenocarcinoma, which represents the most common form of lung cancer in humans (Wynder and Hoffmann, 1994; Hoffmann *et al.*, 2001). Statistics have indicated a rise in the amount of NKK in cigarette smoke correlated with increases in adenocarcinoma of the lung (Wynder and Hoffman, 1994). Additionally, these studies reveal the fact that systemic lung carcinogens such as NNK, are a principal cause in triggering lung cancer in humans.

Epidemiological studies suggested cigar smokers that exhibited an increased risk in lung cancer, (Boffetta *et al.*, 1999; Shapiro *et al.*, 2000). Concurrently, it was established (Hecht and Hoffmann, 2000) that NNK was abundant present in cigar smoke.

Another factor that can lead to the formation of cancerous cells in the lung is the oxidative stress/damage which can also be linked to smoking. Oxidative damage can be responsible for causing DNA lesions, such as 8-oxoguanine. The 8-oxoguanine is a mutation process resulting from the mismatched pairing with adenine (A); causing guanine (G) to pair with thymine (T) and cytosine (C) to pair with adenine which triggers substitutions in the genome (Kanvah *et al.*, 2010; Paz-Elizur *et al.*, 2003).

The most efficient method of reducing lung cancer mortality is prevention. However, since preventing people from starting to smoke is not always successful so the employment of viable cessation techniques is needed. Cigarette smoking is not going to cease in the near future, in fact, there are approximately 1.1 billion smokers worldwide, consuming around six trillion cigarettes per annum (Global Status Report on Noncommunicable Diseases, 2010).

1.3.3 | Causes of Lung Cancer - Passive Smoking

Environmental tobacco smoke, otherwise known as second-hand smoke (SHS), is now widely acknowledged as one of the causes of LC. However, the risk probability is inferior to that of primary smokers. In the context of SHS, epidemiological studies on women with smoking spouses show increases in NNK to levels that were six times higher compared to women who lived with non-smokers (Anderson *et al*., 2001). Similar findings were shown with increased NNK in a group of economically deprived schoolchildren exposed to environmental tobacco smoke (ETS) (Hecht *et al*., 2001). Second-hand smoke can cause numerous health problems in infants and children, including more severe and frequent asthma attacks, respiratory and/or ear infections (Smoking and Tobacco Use, 2016) (Figure 1.5).

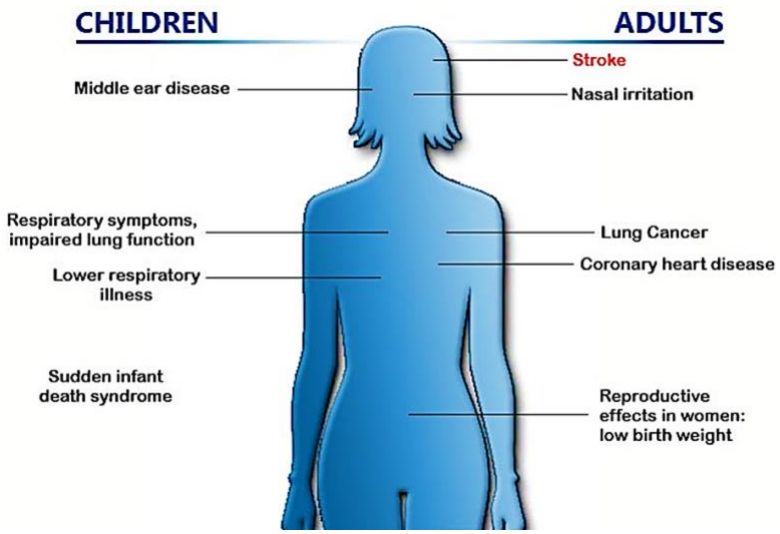


Figure 1.5 | Health consequences in children and adults associated with the exposure to second-hand smoke
It has been proved that there is no risk-free level of passive smoking. Ever since the 1960's (Surgeon General's Report, 1964), it has been established that second-hand smoke (SHS) constitutes one of the causes that triggers lung cancer in humans and it is responsible for the cause of other medical conditions, which are mentioned in this figure. The condition highlighted in red (Stroke), was recently linked to SHS (Surgeon General's Report, 2014). Figure retrieved from Smoking and Tobacco Use website (2016).

Collectively, these observations support the concept of environmental tobacco smoke as a cause of LC, since there is a biochemical link between the uptakes of NNK by non-smokers when exposed to such risk settings.

1.3.4 | Potential Agents that Impact the Progression of Carcinogenesis in Lung Cancer – Socioeconomic and Sociodemographic Causes

In the UK, social status contributes to the increase in percentage of smokers (Bauld, 2006; Cancer Research UK, 2016; ASH, 2001). In addition, a study conducted by Jarvis (2001), showed that 55% in the affluent smokers wished to cease smoking compared to 5% of the less advantaged individuals. Migrant communities that live in the UK fall into this category as well. The figure below (Figure 1.6) illustrates the varied smoking rates among the different communities within the UK population.

Additionally, ASH (2005), confirmed that smokers with a modest income consumed more tobacco than wealthy smokers. Smokers with routine and manual job consume on average 15 cigarettes a day per person, whereas smokers in managerial and professional groups consume 13 cigarettes a day. There is evidence that less affluent smokers spend a seventh of their disposable income on cigarettes, fact that can also determine poorer food choices. They also tended to smoke cigarettes with a higher tar yield, leaving a shorter stub or by drawing harder on the cigarette.

Apart from the association with tobacco consumption, a polluted working environment, a poor diet, low income and less advantaged social backgrounds can affect as the likelihood of developing lung cancer (Rajer *et al*., 2014).

Ethnic Group	Men	Women	Total
White			
White British	27	25	26
Other White	34	26	30
Mixed			
White and Black Caribbean	25	29	28
White and black African	38	26	33
White and Asian	31	33	32
Other Mixed	39	26	31
Asian or Asian British			
Indian	17	4	10
Pakistani	25	6	16
Bangladeshi	45	7	26
Other Asian	26	9	17
Black or Black British			
Caribbean	31	19	24
African	18	5	11
Other Black	19	16	17
Chinese or Other ethic group			
Chinese	34	8	21
Other	33	19	27

Figure 1.6 | Smoking rates of the Black and Minority Ethnic (BME) communities within the UK population
Different smoking percentages of the BME individuals that live in the UK, according to the General Household Survey (2005). 'Other White', 'White' and 'Black African', 'Other Mixed' and 'Bangladeshi', 'Chinese and Other' groups showed significant high percentages compared to the 'White British'. Figure retrieved from ASH, 2015.

1.3.5 | Causes of Lung Cancer – Radon Gas and Asbestos

Exposure to radon gas is, after tobacco smoke, the second main leading cause of LC and the principal environmental cause of LC deaths (Copes, Scott, 2007; EPA, 2009). The risk of developing lung cancer from radon gas exposure is 10 to 20 times higher in smokers when compared to non-smokers (Agency for Toxic Substances & Disease Registry, 2013). It is not just underground miners subjected to increased radon exposure from uranium, tin, silver or coal that develop LC, but households with elevated radon levels have been linked to increased

risk of pulmonary conditions (Field *et al.*, 2000; Agency for Toxic Substances & Disease Registry, 2013).

Asbestosis is a general term attributed to a group of minerals composed of microscopic fibres, which was widely used in construction in the past (NHS, 2014). This group of minerals, if undisturbed, does not present risk factors for individuals' health. However, if the material comprising asbestos is broken, chipped, drilled or allowed in any way to deteriorate, this releases a fine dust composed of asbestos fibres. If the dust containing asbestos fibres is inhaled, it enters the lungs and gradually damaging them over time (NHS, 2014). In the case of prolonged exposure to high numbers of asbestos fibres, asbestosis can develop. Asbestosis is a recognised occupational lung disease, since most cases registered amongst people that work in the Construction Industry. Nowadays, asbestosis became a relatively rare condition, since strict regulations and precautions regarding building construction were put into place, but large amounts of asbestos can still be found in many older buildings.

1.3.6 | Types of Lung Cancer

Lung cancer (or lung carcinoma) classifies into two groups: non-small-cell lung carcinoma (NSCLC) and small-cell lung carcinoma (SCLC). SCLC tends to grow and spread faster than NSCLC (Travis *et al.*, 2004; Healthline, 2015).

1.3.7 | NSCLC

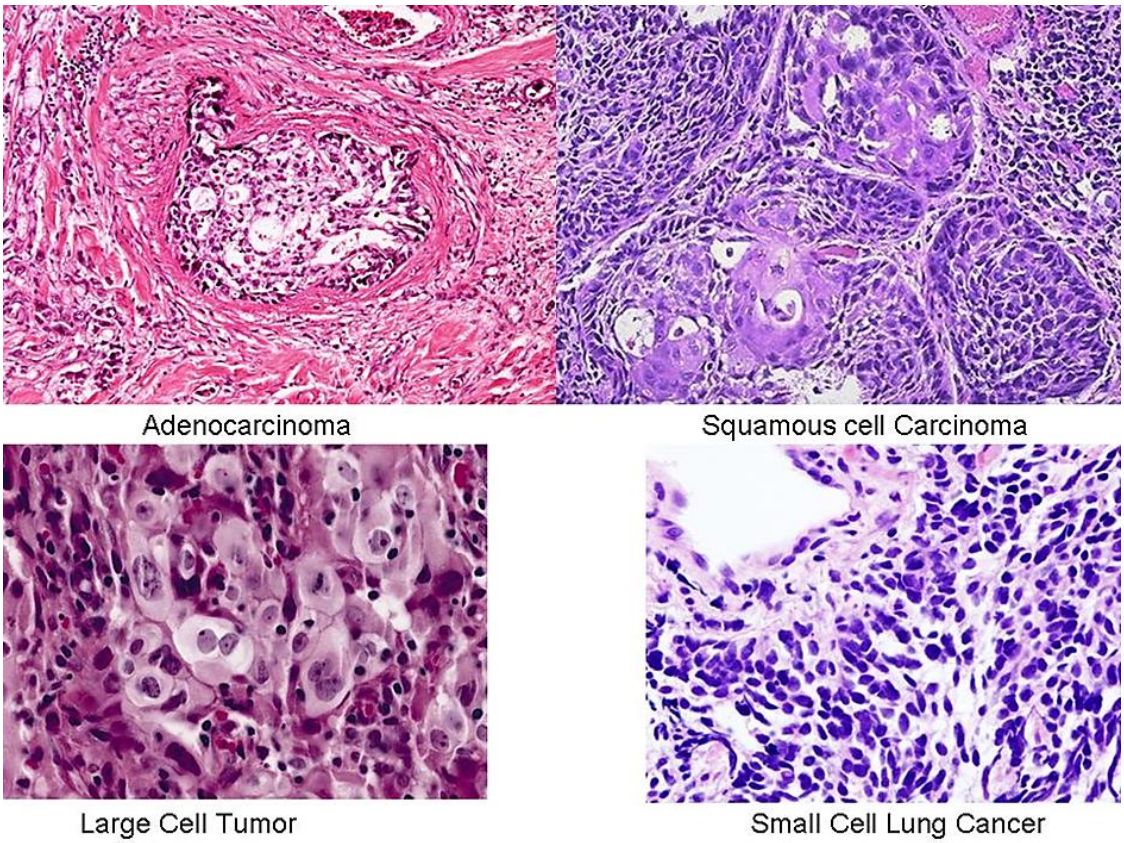


Figure 1.7 | Cytology smears of suspected cases of lung carcinoma with the aim of being utilized as diagnostic material
Cytopathology of non-small-cell lung cancer (NSCLC) including adenocarcinoma, squamous-cell carcinoma and large-cell carcinoma, and small-cell lung cancer (SCLC) courtesy of labroots.com.

Approximatively 85% to 90% of LC cases are NSCLC and are usually classified in three subtypes: adenocarcinoma, squamous-cell carcinoma and large-cell carcinoma laid out in Figure 1.7 (Healthline, 2015).

Of these, adenocarcinoma is a slow-growing type of lung cancer characterised by overproduction of glandular-related molecular products, such as mucin. It is the commonest type to occur in smokers, although it has in non-smokers as well. Squamous-cell carcinoma, the second most common form of lung cancer, normally occurs in the centre of the lung. The

third type of NSCLC, is large-cell carcinoma, as indicated by its name, it is characterised by large cancerous cells with excess cytoplasm and a big nucleus. The large-cell carcinoma is less common, however it can occur anywhere in the lungs and it metastasizes rapidly (Travis *et al.*, 2004).

In order to define how far the cancer has progressed, determine the prognosis and the right form of treatment, the NSCLC stages are assessed via the Tumour, Node, Metastases (TNM) structure. The TNM structure ranges from Stage 0 to IV, and it is intended to assess the primary tumour, the regional nodes and the progression of distant metastasis against established conditions (Travis *et al.*, 2004). Additionally, Stage IV is the most severe and marks the metastasis of lung cancer to other organs or tissues (Figure 1.8).

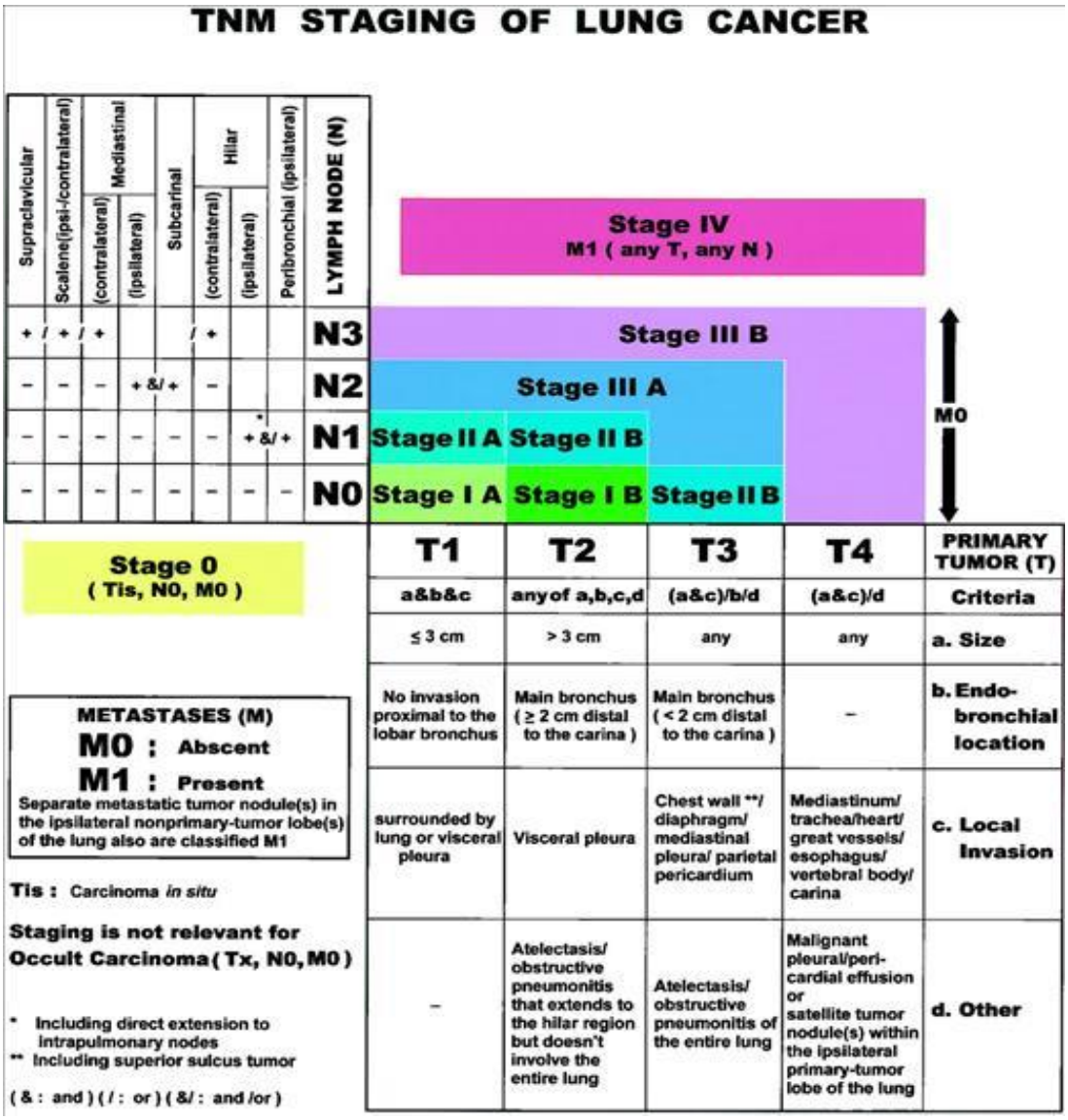


Figure 1.8 | The Tumour, Node, Metastases (TNM) assessment structure
The TNM structure is employed in assessments of the primary tumour, the regional nodes and the progression of distant metastasis in patients diagnosed with NSCLC.

1.3.8 | SCLC

As opposed to non-small-cell lung cancer, small-cell lung cancer is less common and it accounts for 10% out of all lung cancer cases. SCLC tends to commence from the centre of the chest, in the bronchi. It has a typically distinct appearance, characterised by oval rounded small cells or spindle-shaped with minimal cytoplasm and poorly-defined cell edges as shown in

Figure 1.7. Additionally, SCLC differs from NSCLC for the reason that it has a shorter doubling time, higher growth fraction and the tendency to develop earlier metastases in its premature stages. SCLC is staged differently than the NSCLC, as limited or extensive depending on how far the lung cancer has metastasised. This represents the equivalent to the NSCLC TNM staging format (Travis *et al*., 2004).

1.3.9 | Lung Cancer Diagnosis and Treatment

The treatment for lung cancer patients varies with the lung cancer type, stage at diagnosis and treatment. One major issue in the proper treatment is the ability to correctly diagnose the type of lung cancer and its stage, in order to choose the suitable treatment type. Identifying the correct lung cancer diagnosis and/or stage can be very difficult, since the symptoms of the condition resemble other less serious respiratory conditions. Therefore, the pre-clinical diagnosis of lung cancer based on the observed symptoms can be misidentified with those of other pulmonary diseases (Healthline, 2015).

At the moment, there is no evidence of any method aimed to detect lung cancer in asymptomatic patients. Moreover, current methods of early detection of lung cancer can show unclear results or false positives and negatives (Cameron, 2015). This in turn, creates consequences for patients, causing them psychological distress, which leads to increased anxiety caused by the respiratory condition (Byrne *et al*., 2008). However, survival rates in lung cancer are higher for Stages I and II, given that the prognosis is really poor and only 10% of patients remain free of disease (Healthline, 2015). The five-year survival rate registers such a small percentage due to late diagnosis, when the cancer is so advanced that surgical intervention or other clinical involvement proves to be ineffective (Baldwin *et al.*, 2011). Therefore, cheap and effective screens for lung cancer are needed.

1.3.10 | Screening for Lung Cancer

Currently, screening techniques for lung cancer have proved to identify patients at high-risk of developing the disease by using three methods: physical examination, spiral computerized tomography (CT) scan and biopsy assessment.

1.3.11 | Screening for Lung Cancer – Medical Examination

The medical examination focuses primarily on signs of cough (chronic or recurrent) or coughing up phlegm that contains blood, chest pain and shortness of breath, wheezing or other potential symptoms that indicate the presence of LC in a patient. The patient's history of smoking and a chest X-ray typically in the form of a CT scan, also have to be included in the physical examination (OnHealth, 2016).

1.3.12 | Screening for Lung Cancer – Spiral CT Examination

This technique of CT scanning constructs detailed images of the body internally. Hence, when coupled to an X-ray machine it generates 3D images of the patient's internal organs where potential cancerous tumours may be revealed (OnHealth, 2016).

Study trials based on the efficiency of CT scans, carried in different countries, have given inconclusive results. A Japanese research study conducted by Nawa *et al.* (2012) on low-dose CT (LDCT) scans concluded that out of a total of 61,914 CT scans, 25,385 patients were screened for lung cancer and only 210 individuals were identified with primary lung cancer. Within these 210 patients, 203 underwent surgery with an outcome of 90% success regarding the five-year survival rate for all patients. Consequently, CT scans constitute an effective method of identifying lung cancers. However, the cost of achieving computed tomography

scans to identify merely 0.83% of individuals suffering from lung cancer might prevent its implementation in a national screening procedure, due to cost.

Similarly, an Italian study (Infante *et al.*, 2009), confirms that spiral CT scans may be over-estimated. In addition, spiral CT scans identified a large number of Stage I lung cancer patients in comparison to its randomised control group. However, in the three-year follow-up assessment it was confirmed that there has been minor variation between the survival rates of lung cancer cases identified through spiral computed tomography and those detected using conventional means.

At best, the screening process based on spiral computed tomography scan detects approximatively 30% of lung cancers, which leaves the majority of 70% cases undetected (OnHealth, 2016). If the potential of screening through CT scans does not meet the expected outcome of detecting lung cancer at early stages, the next steps are biopsy sample examination or ultimately, surgery. For this reason, new screening methods for lung cancer are sorely needed.

1.4 | Pneumonia

A general consensus has defined pneumonia (or pneumonitis) as a group of acute (co)infections of the lung parenchyma (lung sections such as alveoli or the air sacs involved in gas transfer) that is caused by various pathogens. It is usually considered separately to bronchiolitis due to its well-defined symptoms (Mackenzie, 2016).

Pneumonia is associated with an inflammation of the lungs, specifically in the alveoli that fill up with fluid, causing difficulties in their normal function. Due to the intense stress received from the infection, the body sends neutrophils and macrophages to the contaminated site to

exterminate the foreign body; however, their accumulation can impose other difficulties for the lungs to deliver oxygen into the bloodstream (British Lung Foundation, 2012).

The issue of defining and classifying pneumonia has been problematic often resulting in experimental or unnecessary use of antimicrobials that cause life-threatening pathogen resistance or contributing to heterogeneous pathologies and biased classification amongst clinical phenotyping (Mackenzie, 2016). To reduce the threats of the widespread empirical use of antibiotics and maximizing the validity of clinical studies, Levinson (2001), suggested that the selection of individuals for study groups should be less generalised. Therefore, patients should be selected that share similar clinical phenotypes and to increase the effectiveness and internal validity of the clinical study. For instance, a research study based on the evaluation of different antibacterial therapies in patients suffering from pneumonia, should be based only on patients with a proven bacterial cause; as opposed to using patients with viral pneumonia.

If the methods to determine the aetiology of pneumonia are improved and the clinical studies focus more on subjective methods of classification, perhaps research questions regarding the diagnosis and management of pneumonia would deliver clearer answers.

1.4.1 | Pneumonia Statistics

Approximately 220,000 people are diagnosed with pneumonia each year. In 2012, a higher proportion of individuals were identified with pneumonia, particularly in the East Midlands and the North West of England (due to industrialization) than in the UK overall. On the other hand, Scotland, Wales and the other regions of England, indicated lower proportions as displayed in Table 1.7 (British Lung Foundation, 2012).

With regards to the gender proportions, the British Lung Foundation (2012) indicated an increase in pneumonia for both sexes of 329 males and 361 females per 100,000 in 2012, down

from the period of 2004 when the maximum rate was of 292 males and 321 females per 100,000 people. The high prevailing cases of pneumonia in women may be caused by a viral strain of the disease, which tends to be more severe in pregnant women and in those suffering from heart diseases, where female cases predominate (Dallas, 2013; Diseases and Conditions, 2016).

In principle, pneumonia can affect people in any age group, although the most frequent cases are registered in children younger than five years old and in the elderly people with ages over 60 (British Lung Foundation, 2012; WebMD, 2016). Smokers or those that suffer of other respiratory diseases, heart, kidney or liver diseases are susceptible as well (WebMD, 2016). Pneumonia cases are more common during autumn and winter and are influenced by the cold changes in temperature or damp weather conditions, which provide the perfect environment for the development of viruses and bacteria.

Table 1.7 | Total number of individuals diagnosed with pneumonia for every 100,000, over the period of 2008-2012

The majority of pneumonia cases prevail in regions of England such as: West and East Midlands, North East, North West and in Yorkshire and Humber (colour coordinated arrows marked according to every region mentioned). Scotland and Wales, as well as other regions of England revealed lesser cases in comparison to the previous regions (British Lung Foundation, 2012).

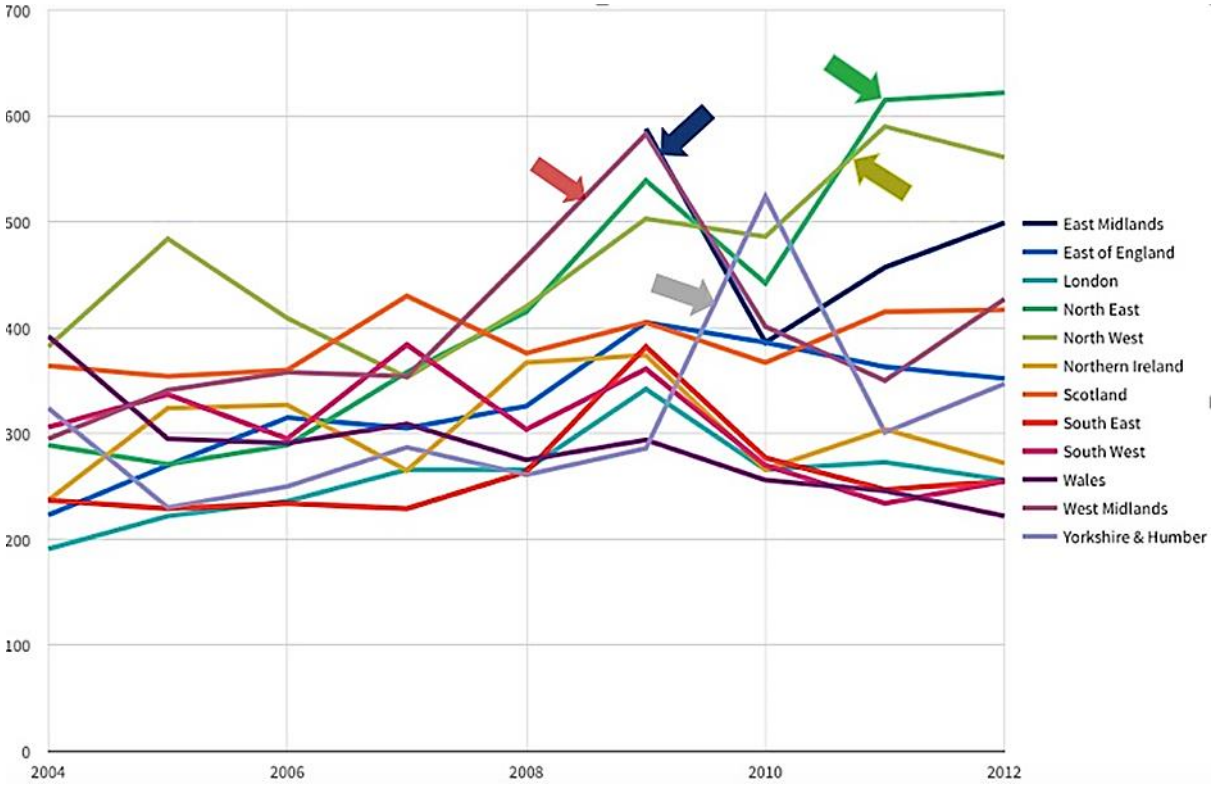
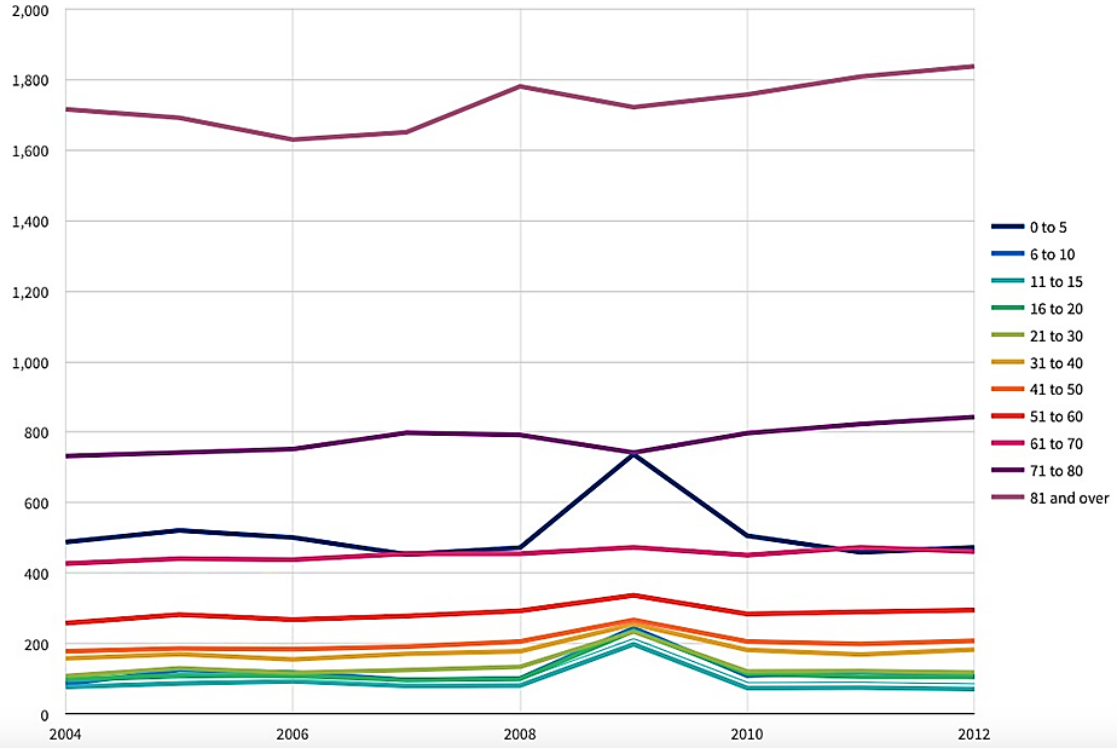


Table 1.8 illustrates the most predisposed age groups of people to develop pneumonia during the period of 2008-2012 with a total of 473 new cases per 100,000 amongst young children aged 0-5, 843 for every 100,000 in adults between 71-80 years old and 1,838 per 100,000 in old individuals aged 81 and above (British Lung Foundation, 2012).

As a disease, pneumonia is not contagious, although a healthy person becomes susceptible after coming in contact with the pathogens of an ill individual by breathing in droplets from coughs or sneezes, or even getting in contact with the environment that contains the pneumonitis viruses or bacteria. For this reason, pneumonia mortality rates indicated this disease as one of the three major respiratory causes of death in the UK, after lung cancer and COPD (Figure 1.1). Additionally, during the period of 2001-2010 approximately 214 individuals per every million died from pneumonia in the UK, leaving the country just outside the top of 20 countries with deaths. The only European countries that indicated higher mortality percentages in pneumonia than the UK were Slovakia and Romania (British Lung Foundation, 2012).

Table 1.8 | Age group classification of people in the UK that developed pneumonia over the period of 2008-2012

The prevailing cases of people that died from pneumonia during 2008-2012. Most predisposed age groups of people to develop pneumonia are children below age 5 and old people aged 60 or above (British Lung Foundation, 2012).



1.4.2 | Causes of Pneumonia

More than 30 pathogens including bacteria, viruses and fungi have been found to cause pneumonia. The latency period of pneumonia is relatively short, the foreign organism will rapidly cause inflammation of the lung alveoli triggering the production of fluid and pus, as the person will become visibly ill experiencing chills, fever, coughing and difficulty in breathing (Pneumonia Symptoms and Causes, 2013).

Pneumonia is not considered a contagious disease, although its causative pathogens are extremely dangerous because they are found in the commonest places or in environments that people frequent on a daily basis. For instance, children that attend school or day-care centres are more predisposed to catching viruses from each other, which increases their chances of developing viral pneumonia. Additionally, the viral strains causing pneumonia are more transmissible and likely to spread from one another in comparison to its other bacterial and fungal sources. An easily transmissible bacterial strain causing pneumonia is *Legionella pneumophila*. This species is transmissible only in specific environments for example, in certain contaminated air conditioning system. Studies have also indicated that inhaling droplets from spas, fountains or whirlpools may be linked to contamination with the bacterial strain of pneumonia as well (Pneumonia Symptoms and Causes, 2013).

1.4.3 | Types of Pneumonia – Bacterial

The most recognized species causing bacterial pneumonia include: *Streptococcus pneumonia* (pneumococcal pneumonia) extremely common in adults; *Legionella pneumophila* (Legionnaire's disease); *Mycoplasma* pneumonia which affects young adults with occupations related to crowded areas such as homeless shelters, schools or prisons; *Chlamydophila*

pneumonia which is milder and common in people over the age of 60; *Haemophilus influenzae* type B usually seen in children with ages below five years old.

The majority of bacterial pneumonias are pyogenic (increased mucus production), therefore symptoms as cough with phlegm (usually thick yellow) production represents a typical sign of bacterial pneumonia. From this, a series of other symptoms are triggered, such as pain when coughing or inhaling deeply, breathlessness and exhaustion, high fever and loss of appetite (Pneumonia Symptoms and Causes, 2013; WebMed, 2016).

More advanced studies (Marrie *et al*., 2000; Steel *et al*., 2013) have concluded that community-acquired pneumonia (CAP) caused by *Streptococcus pneumoniae* (or the pneumococcus) is the leading cause of morbidity and mortality linked to pneumonia. Regardless of the efforts on implementing the pneumococcal polyvalent vaccine as a national strategy of immunization for people at high-risk, *S. pneumoniae* is still the commonest cause of CAP. Hospital-acquired pneumonia, fits under the subdivision of CAP as it particularly refers to the transmission of pneumonia to a patient, whilst being admitted for a different condition. *S. pneumoniae* has also been linked to other medical conditions such as septic shock, requirement for vasopressors, mechanical ventilation and consequently for an extended period of hospitalization in ICU (Rice *et al*., 2012). Moreover, serotype 3 was identified as a cause in triggering the invasive pneumococcal disease (IPD) and also associated with septic shock. In the attempt of understanding the microbiology of the pneumococcal infection, studies have focused on its serotypes severity with the aim of creating novel types of pneumococcal conjugate vaccines (Ahl *et al*., 2013).

CAP has become one of the most dangerous infectious diseases in many parts of the world (Brown, 2012; Fauci and Morens, 2012; File and Marrie, 2010; Welte *et al*., 2012). The risks of CAP infections increase with the aging populations, as in Europe (Blasi *et al*., 2012).

Several clinical studies (Ho *et al*., 2006; Fuller *et al*., 2005) have acknowledged that the pneumococcal bacterial load is linked to the severity of infection and has predictive value. Consequently, repeated measurements of the bacterial load based on reliable assays could aid in a more accurate technique of monitoring the treatment progress of CAP.

Advanced clinical studies based on the biomarkers capacity to identify inflammatory markers in patients with CAP, were associated with WBC count, acute phase reactants such as C-reactive protein (CRP) also identified in patients suffering from COPD, cytokines such as interleukin-1 (IL-1) and tumour necrosis factor-α (TNF-α), cortisol and stress hormones (Kolditz *et al*., 2013; Pereira *et al*., 2012). Of these, the ones that have received particular attention and have been mentioned in many research studies are the C-reactive protein (CRP) and procalcitonin (PCT). Additionally, CRP and PCT have been reported as useful tools of assessment in determining the severity of CAP and if utilised in combination with other scoring evaluation systems such as CURB-65, could increase the diagnosis accuracy not only amongst the different strains of pneumonia, but for other respiratory diseases that are characterized by similar symptoms (Pereira *et al*., 2012; Lippi *et al*., 2011; Steel *et al*., 2013).

1.4.4 | Types of Pneumonia – Viral

Viral pneumonia is usually identified in children with ages below five and accounts for approximately a third of all pneumonia types. Although the viral strain of pneumonia is less severe and mends in a few weeks, it can leave the individual more susceptible to bacterial pneumonia. The symptoms of viral pneumonia usually resemble influenza as the person can

experience a sore throat, loss of appetite, fever or muscle pain. One major symptom difference that differentiates viral pneumonia from the bacterial type is the nature of the cough that is dry, hence, lacks mucus production. Viral pneumonia, usually caused by the respiratory syncytial virus (RSV) and occasionally it can be triggered by influenza (Pneumonia Symptoms and Causes, 2013).

In 2003, a new type of viral pneumonia, severe acute respiratory syndrome (SARS) was identified, which was triggered by the same type of viruses that produce the common cold. About 8,000 cases of SARS were reported at the time, but since then the incidences have become less common (Pneumonia Symptoms and Causes, 2013).

1.4.5 | Types of Pneumonia – Fungal

The least common type of pneumonia and rarest in the UK, is the fungal one. According to the literature, there are three species of fungi (living in soil) that have been responsible for causing pneumonia: *Coccidioidomycosis* (Galigiani, 1993), found in South-West of America; *Histoplasmosis (*Deepe Jr, 2018*)* and *Cryptococcus* (Soltani *et al*., 2013) found in bird droppings.

Fungal pneumonia appears to often be related to the degree of weakness of the immune system and any virulent pathogenic mechanism(s). For example, the *Pneumocystis jirovecii* fungus has been found to cause pneumonia in patients suffering from human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) (Pneumonia Symptoms and Causes, 2013).

1.4.6 | Treatment and Management of Pneumonia

Pneumonia diagnoses usually involve physical examination, chest X-rays, cytology of a phlegm sample, a blood test of the patient and investigation of the medical history (Pneumonia Symptoms and Causes, 2013).

For patients with milder cases of bacterial pneumonia that present a low probability of worsening are usually treated with a course of antibiotics along with a rest at home and to keeping hydrated at all times.

Patients that suffer from more severe cases of bacterial pneumonia are often infected with strains exhibiting antibiotic resistance. This issue has been most often encountered in patients CAP caused by *S. pneumoniae*. In such situations, the virulence of the pneumococcus coupled with the inflammatory responses that it elicits, contribute to the disease invasiveness and symptom development.

Prior to any action being taken or drug administration, every patient suspected with the pneumococcal infection, is evaluated using the Pneumonia Severity Index (PSI) or CURB-65 to indicate the seriousness of the illness (Niederman, 2007; Kolditz *et al*., 2013). The PSI aims to assist in identifying patients that suffer from the bacterial pathogen and can still be managed at home. It is based on 20 variables such as patient age, gender, existence of comorbid conditions, vital signs abnormalities and several analyses of laboratory and radiographic parameters (Feldman, 2007). The CURB-65 on the other hand, was designed to document more severely ill patients, with particular interest for those admitted in the intensive care unit (ICU) (Pereira *et al*., 2012). CURB-65 contains only five variables including the levels of serum urea (>7 mmol/L), the respiratory rate (≥30 breaths/minute, the low blood pressure (systolic <90

mmHg or diastolic ≤60 mmHg), the age threshold ≥65 years and the confusion level (Feldman, 2007; Steel *et al*., 2013).

Furthermore, in terms of CAP this can be complicated both seasonal and pandemic influenza (Steel *et al*., 2013). Indeed, co-infections with influenza strain H1N1 have been detected in patients who were initially admitted to the hospital only with CAP (Cilloniz *et al*., 2012). Pneumococcus has proven to be the most prevailing bacterial species in co-infections (62% of cases) (Rice *et al*., 2012).

Due to its life-threatening nature, treatment of CAP and more specifically the pneumococcal disease also involves adjunctive anti-inflammatory therapies in order to suppress the pathogen-activated inflammatory damage to the lungs (Steel *et al*., 2013). Three anti-inflammatory agents may be employed: macrolides, corticosteroids, and statins (Feldman and Anderson, 2009; Steel *et al*., 2013).

Macrolides possess both pathogen- and host-directed anti-inflammatory properties. These can inhibit bacterial protein synthesis and secondarily, unrelated to their antibacterial activity, macrolides control the recruitment and mobilisation of neutrophils by reducing the production of neutrophil-mobilising chemotactic cytokines interleukin-8 (IL-8), interleukin-17 (IL-17) and tumour necrosis factor (TNF). Macrolides have been proved to have a beneficial therapeutic role with respect to pathogen- and host-directed anti-inflammatory activities in a few chronic respiratory diseases as well, such as cystic fibrosis and COPD (Steel *et al*., 2012).

Corticosteroids (CS) comprise a broad range of anti-inflammatory agents, although their adjunctive role in patients suffering from severe CAP remains to be confirmed (Steel *et al*., 2013). A retrospective study on patients with severe CAP that received systemic CS for a period of three days, reported an amelioration of their medical state (Mikami *et al*., 2007).

Subsequently, Meijvis *et al*. (2011), in a large randomized, double-blinded, placebo-controlled study based on low dose systemic administration of dexamethasone (corticosteroid medication) to non-immunocompromised patients, demonstrated a significant decrease of the hospitalization period. The effectiveness of corticosteroids in patients with CAP appears to be influenced and determined by the pathogen type with, *S. pneumoniae* most responsive to its effects (Meijvis *et al*., 2011).

Statins refer to a class of pharmacological inhibitors of the enzyme 3-hydroxy-3-methylglutarylmcoenzyme A reductase. This enzyme has cholesterol-lowering properties; therefore, it is utilised to regulate hypercholesterolemia to reduce the risk of cardiovascular diseases and stroke in people. It has been established that this statin also possesses important anti-inflammatory capacities (Steel *et al*., 2013). Several systematic studies in the field (Corrales-Medina and Musher, 2013), have reported improved outcomes and reduction in mortality by 50% in patients suffering of pneumococcal CAP after the administration of statins (Steel *et al*., 2013). This suggests that statins protect against the invasiveness and pro-inflammatory activities in diseases caused by *S. pneumoniae.*

To conclude, there is a clear need for improved strategies of determining the prognostic indicators of pneumonia and its microbial causes in order to establish more accurate diagnoses. This in fact, could reduce the excessive use of antibiotics, and hence, moderate the bacterial resistance as well, providing a less negative impact on healthcare costs and mortality prevalence.

1.5 | Asthma

Asthma has been defined as a common, long-term respiratory disease that entails ongoing management depending on its severity. It can affect all age groups but most often, asthma starts during childhood (British Lung Foundation, 2012; World Health Organization, 2012.).

Asthma is characterized by inflammation of the lung airways that can easily become irritated due to a series of factors such as cold air, exercise, stress, inhalation of certain substances (i.e. smoke), pollution or pollen (British Lung Foundation, 2012). The usual symptoms are shortness of breath and the feeling of a "tight chest", cough and wheezing. During an asthma attack, the feeling of "tight chest" is actually caused by the swelling of the lining of the air passages which restricts the airways and as a consequence, the air flow that circulates in and out of the lungs is reduced (World Health Organization, 2012). Asthma symptoms tend to become more severe at night, particularly in the case of those that do not manage the respiratory condition properly. This in fact, is a consequence of the body's natural process of suppressing inflammation (of the airways as well) that tends to switch off or decelerate while an individual is asleep (Asthma UK, 2015).

Asthma symptoms can range from mild, which means that the person affected will experience it occasionally with very few symptoms, to severe asthma, in ~ 5% of cases, and these require medical assistance and support in order to manage the symptoms. Unfortunately, asthma attacks are mainly caused by people's negligence in either taking medicines or failing to follow the correct medical treatment (Asthma UK, 2015).

Based on the factors that trigger the condition, asthma has been categorised into two types: allergic and non-allergic asthma. Allergic asthma is usually caused by allergens such as pollen,

animal hair or dust mites. Conversely, the common factors that cause the non-allergic asthma symptoms include colds and flu produced by the cold weather, intense exercise, or irritants as cigarette smoke and car exhaust fumes (Asthma UK, 2015).

1.5.1 | Statistics of Asthma

In the UK, around eight million people have been identified with asthma over the period of 2004-2012, according to the British Lung Foundation (2012). However, statistical studies indicate that asthma may be over-diagnosed and new methods of identification are needed in order to differentiate the people that are living with asthma and are still receiving medication for it, from the ones that were diagnosed during childhood and grew out of it, or ceased their treatment despite the continuous symptoms occurrence.

As shown in Table 1.9, between 2008 and 2012 the number of people for every 100,000 that have been identified with asthma indicated high rates in all parts of the UK. Although the more cases were seen in the East Midlands, the East of England as well as the North West and South West in comparison with the other parts of the UK.

Considering the prevalence of asthma, it has been suggested that women are more affected by the disease than men (British Lung Foundation, 2012). Between 2008-2012 statistical tests indicated 284 per 100,000 new cases of asthma in women, as opposed to 261 for every 100,000 in men. One factor that influences this could be the rate of obesity that also appears to be increasing faster in women than in men (Chen *et al*., 2002; Nystad *et al*., 2004).

Nystad *et al*. (2004), showed that there is a link between the BMI (weight (kg)/height $(m)^2$) of a person and the probability of developing asthma symptoms in a longitudinal assessment of 135,000 Norwegians with ages ranging from 14 to 60 years old. Based on the Cox proportional hazards regression, study participants were evaluated depending on their smoking status,

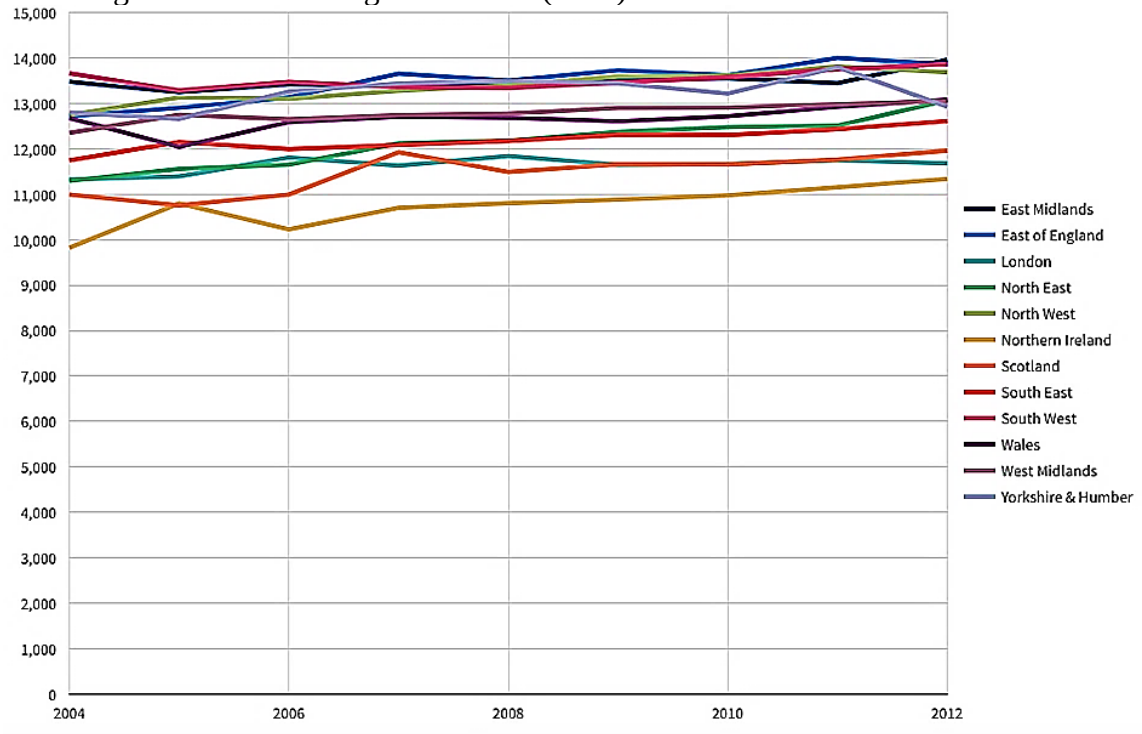
education and physical activity with the aim of estimating the relative risk of asthma in relation to these variables. Contrasts between men and women with a BMI of less than 25 to the ones recorded as overweight (BMI = 25-29) and obese (BMI ≥30), showed relative asthma risks for the latter two categories. The final outcomes revealed proportions of 1.27 with a 95% Confidence Interval (CI) of 1.13 in men and 1.43 in women with BMIs of less than 25; 1.30 in overweight people (95% CI: 1.17 men, 1.45 women) and 1.99 (95% CI of 1.67 men, 2.37 women) in people recorded as obese (BMI ≥30). The investigation findings confirmed that the risks of asthma increase, as the BMI gets higher with an average of BMI 20 in men and 22 in women across the proposed strata of smoking status, education and physical activity.

Consequently, obesity represents a risk factor in the development of asthma in both men and women and can be influenced by the quality of life-style as well.

Regarding the mortality in people with asthma in the UK, the figures are lower, compared to other respiratory diseases suggesting that asthma is easier to manage. This stated around 1,200 people a year die of asthma (British Lung Foundation, 2012) which should raise more awareness regarding the managing techniques.

Table 1.9 | Number of individuals per 100,000 identified with asthma in the UK over the period of 2008-2012

The prevailing rate of people identified with asthma tends to be higher in specific parts of the UK such as the East Midlands, the East of England, the North West and the South West, with lower cases indicated in Northern Ireland, Scotland, North East and London according to the British Lung Foundation (2012).



1.5.2 | Management of Asthma

Concerning the guidelines on asthma management, there is a constant effort to improving the flexibility of the medication in order to make it more appropriate according to the severity type and responses to medical intervention in every individual. This patient specific response to treatment has been demonstrated in a randomised double-blind, parallel-group test has considered patients with asthma that were still experiencing symptoms, regardless of the administered dose of inhaled corticosteroids (Greening *et al.*, 1994). The asthmatic patients were divided into three groups according to the medication received: the first group were given inhaled beclomethasone dipropionate (BDP), the second group salmeterol xinafoate and BDP, and lastly, the third group a higher dose of BDP twice a day. The peak expiratory flows (PEFs)

of every patient were measured three times a day, every day for six months. Results indicated an improvement in the morning measurements of the PEF in patients that received salmeterol/BDP and the ones that received a higher dose of BDP. Although, the increased peak expiratory flow at all time points was observed only in those given salmeterol/BDP with the occasional rescue use of the bronchodilator salbutamol (Greening *et al*., 1994).

The Global Initiative for Asthma (GINA) formed in 1993, disseminates information regarding the management of patients with asthma, based on the published clinical literature. The report "A Global Strategy for Asthma Management and Prevention" published in 1995 and reconsidered in 2006, to inform numerous national guidelines with respect to asthma control and management as the focus for the treatment choice. Another two factors emphasised in GINA, were the significance of the patient – caregiver relationship and guided self-management (Bateman *et al.*, 2008).

However, the Asthma Insights and Reality in Europe (AIRE) survey conducted by Vermeire *et al*. (2002) assessed the prevalence of asthma control in patients from Western European countries as France, Germany and the UK, suggested that the GINA guidelines did not meet its objectives. According to Vermeire *et al*. (2002), the symptoms in both adults and children suffering from asthma were poorly controlled due to poor implementation of disease management plans by more than 50% of adults and 61% children in all countries. Follow-up visits never occurred unless asthma symptoms worsened or for the required prescription of inhaled corticosteroids or short-acting bronchodilators.

That is why there is a constant need for further efforts to thoroughly employ the importance of considering the asthma management guidelines in order to improve asthma control worldwide.

1.6 | Idiopathic Pulmonary Fibrosis (IPF)

Idiopathic pulmonary fibrosis (IPF) has been defined as a progressive lung condition that deteriorates the lungs with time. Pulmonary fibrosis is characterised by the development of scar tissue in the lungs, causing them to become rigid and lose their elasticity when expanding, which consequently leads to serious breathing problems such as breathlessness of exertion, increased cough and dyspnoea as well as finger clubbing. The idiopathic term indicates the unknown aetiology from which IPF arises, making it difficult to predict its progress since it occurs spontaneously. Even when the condition has been stabilized, it has been shown that people can get sudden outbreaks of acute exacerbations. The diagnosis of IPF can be achieved through a chest X-ray sometimes followed by a CT scan for a more in-depth image of the lungs, breathing test measurements or a bronchoscopy which allows the specialist to view inside the lungs and obtain a small sample of the lung tissue for biopsy (British Lung Foundation, 2012). Idiopathic pulmonary fibrosis is an irreversible respiratory condition, which affects the alveoli (air sacs) by limiting the volume of oxygenated blood, which is intended to supply the lungs with (Figure 1.9).

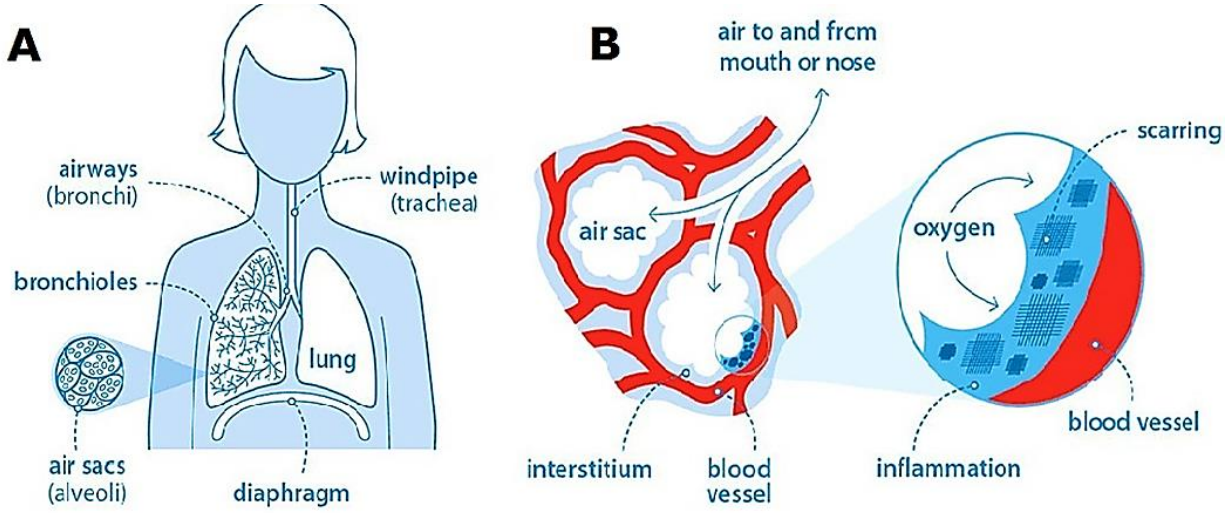


Figure 1.9 | Illustration of the brief anatomy of the lungs and the development of idiopathic pulmonary fibrosis (IPF)
(A) The air inhaled through the mouth and nose is drawn into the trachea, which divides into two bronchi that are linked to the lungs. The air then travels down the bronchi and into smaller airways called bronchioles. Bronchioles are formed of small air sacs called alveoli, which move the oxygen across small blood vessels and finally into the blood. **(B)** Following, the alveoli exchange the waste gas and carbon dioxide from the blood with the inhaled oxygenated air. In the case of IPF, the scarring of the alveoli can reduce the volume of inhaled air that is needed to oxygenate the blood and consequently causes the symptom of breathlessness (dyspnoea) in a person (British Lung Foundation, 2017).

1.6.1 | Statistics in Idiopathic Pulmonary Fibrosis

According to the British Lung Foundation (2012), approximately 6,000 individuals are diagnosed with IPF every year in the UK with a rate of 85% of diagnoses recorded in people over 70, mostly prevailing in men. The higher incidence of IPF in men compared to women (Nalysnyk *et al*., 2012) is a consequence of men being more exposed to environmental factors linked to working labour (British Lung Foundation, 2012). A history of occupational and environmental fibrogenic exposures as inhalation of mineral dusts including coal, asbestos, silica, beryllium (found in mineral rocks), hard metal (cobalt) or a prolonged history of ingestion of pharmaceutical substances such as nitrofurantoin and amiodarone may also constitute factors that cause IPF (Khalil *et al*., 2007).

1.6.2 | Causes of Idiopathic Pulmonary Fibrosis

The body generates fibrosis in response to factors that damage the lungs (Panos *et al*., 1990). The histological equivalent to IPF is interstitial pneumonia, both of which are characterised by scarring of the lung (fibrosis) combined with healthy alveolar tissue. This is caused by repetitive episodes of lung injury, which finalise in abnormal lesion recovery responses. The hypothesized types of injury are believed to be immunological, microbial or chemical, also including the aspirated gastroesophageal reflux (GER). Consequently, some individuals are predisposed to IPF, through repeated steps of microaspiration of gastric refluxate (Fahim *et al*., 2010). Certain viruses have been also identified to contribute to the development of IPF, including the Epstein-Barr virus (EBV) responsible for causing glandular fever (Sitki-Green *et al*., 2004), the oral virus herpes simplex (HSV-1) and hepatitis C virus (HCV) (British Lung Foundation, 2012).

A well-defined classification criterion for IPF and a possible gold standard test may provide a clearer understanding of the pathogenesis of the disease and therefore, provide a better understanding that IPF encompasses.

1.6.3 | Treatment and Management for Idiopathic Pulmonary Fibrosis

Due to the irreversible nature of idiopathic pulmonary fibrosis, current treatments only achieve a slower rate of the lung scarring, unable to prevent it. Currently in the UK, IPF treatments have increased the survival rate from untreated situations where life expectancy in individuals diagnosed with IPF was lower than three years due (British Lung Foundation, 2012).

Patients may receive treatment depending on the IPF causes and severity. One approach is pulmonary rehabilitation, which is usually tailored according to every patient and may or may not consist of pharmacologic intervention.

Pulmonary rehabilitation is a maintenance exercise programme for patients who remain symptomatic despite the treatment or suffer from a pulmonary impairment (Respiratory Care, 2002) with the aim of teaching them how to handle their respiratory condition (British Lung Foundation, 2012).

In more severe cases of IPF, patients require the home oxygen therapy. The oxygen treatment is based on a portable cylinder, which can be easily operated even from home. The aim of the home oxygen therapy is to increase the oxygen level in the lungs and thus, in the bloodstream in order to avoid heart or brain damage. Additionally, N-acetyl-cysteine (NAC) has proven to be of great help in reducing the mucus secretion in the lungs (British Lung Foundation, 2012).

The current pharmacologic treatments licensed to be used in patients suffering from IPF are pirfenidone and nintedanib. These two medicines act by decelerating the progress of the fibrotic tissue in the lungs (British Lung Foundation, 2012) most likely through their action on vascular endothelial growth factor (VEGFR) and fibroblast growth factor (FGFR) receptor.

Recent studies in the field (Flynn *et al.*, 2015) have approached the challenging problem of IPF management via biomarkers that could provide early prediction of the disease. Biomarker candidates found in culprit cells or genes from biofluids such as saliva, blood or urine have provided a better insight of the IPF pathogenesis. Moreover, IPF biomarkers possess the ability of revealing the natural course of the disease without influencing it, as opposed to medical treatment.

A group of effective indicators of IPF fibrotic are the metalloproteinases (MMPs). These zinc-dependent enzymes are actively involved in hydrolysing components of the extracellular matrix (ECM) and they represent an essential key factor in biological processes as organogenesis or inflammation and in diseases such as arthritis, cancer or tissue ulcerations (Visse, Nagase,

2003; Flynn *et al*., 2015). Metalloproteinases are the enzymes responsible for the degradation of the ECM. Therefore, dysregulations in the composition, structure, rigidity and abundance of the ECM caused by the MMPs, may be the factors that contribute to the development of fibrosis and invasive cancer, or other pathological conditions (Bonnans *et al*., 2014). MPP have also been shown to play an important role in the imbalance of the collagen turnover in fibrotic lung diseases. An evaluation study of the MMP levels has identified that particular matrix metalloproteinases such as MMP3, MMP7, MMP8 and MMP9 were more increased in the bronchoalveolar lavage fluid (BALF) of patients with IPF compared to the healthy controls (McKeown *et al*., 2009). Moreover, MM7 has been indicated to exhibit a higher degree of gene and protein expression in the alveolar and bronchiolar cells of fibrotic lungs (Zuo *et al*., 2002). Similarly, MMP8 and MMP9 were linked to the rapid decline of the lung function and high incidence in mortality (Craig *et al*., 2014).

Overall, MMP7 has proven to hold an excellent clinical utility as a potential biomarker in monitoring the development of idiopathic pulmonary fibrosis, and in combination with other biomarkers of IPF, MMPs may have a role in the prognosis and disease progression of respiratory conditions with fibrotic features.

1.7 | Bronchiectasis

Bronchiectasis is an irreversible respiratory disease characterised specifically by inflammation of the airways (bronchi) and build-up of mucus, which eventually become infected due to bacterial accumulation. In cases where bacteria subsist in the airway tubes, these become inflamed and consequently result in respiratory impairment (British Lung Foundation, 2012).

1.7.1 | Bronchiectasis Statistics

According to the British Lung Foundation, in 2012, approximately 210,000 individuals in the UK were suffering from bronchiectasis. However, the high incidence of bronchiectasis requires more thorough statistical measurements to confirm the existing information and to establish whether the disease is becoming more common or is being identified more accurately. Bronchiectasis is most common in the West Midlands more than any other region in the UK, with the lowest rate identified in the South East of England (British Lung Foundation, 2012) (Table 1.10).

In the recent years, there appears to be an increase of 33 per 100,000 of newly diagnosed individuals with bronchiectasis in women compared to men. Most people affected by bronchiectasis are aged 70 or over which accounts for 60% of all patients. Bronchiectasis accounts for 0.3% of all deaths in the UK and 1.4% of deaths caused by lung disease (British Lung Foundation, 2012).

1.7.2 | Causes of Bronchiectasis

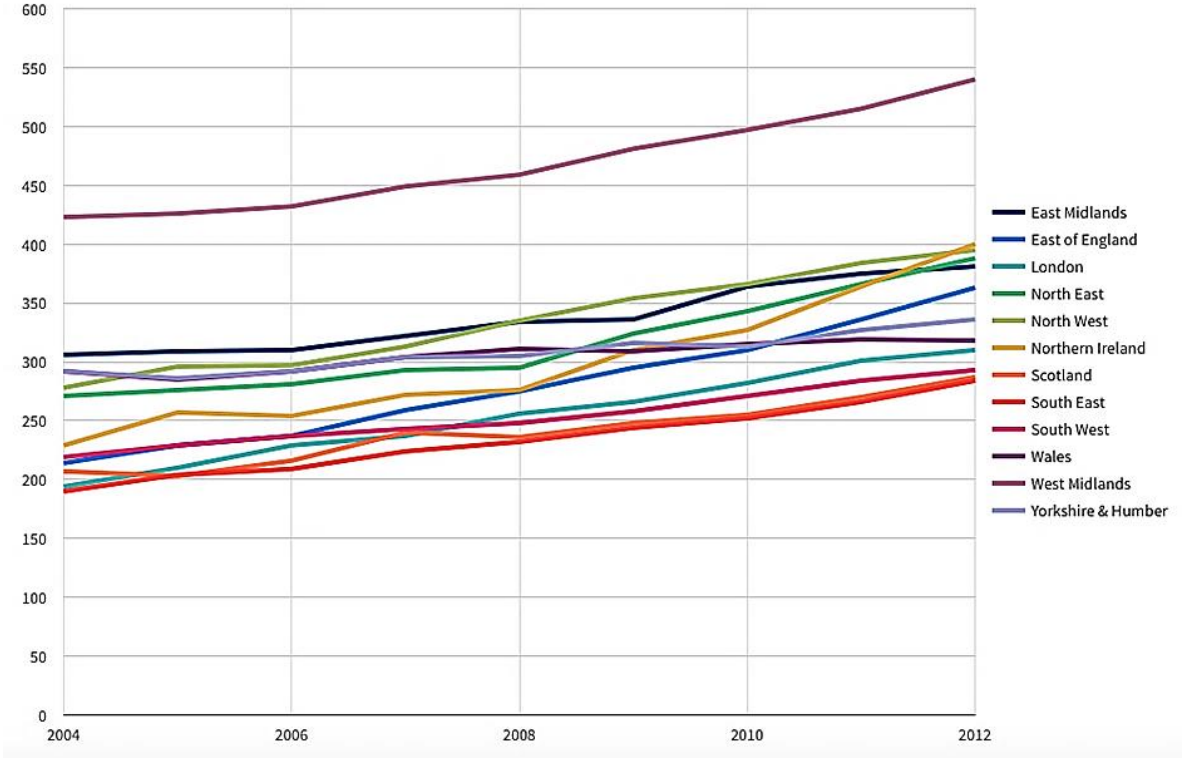
Similarly to IPF, the causes that trigger bronchiectasis remain unknown, yet the consequences always consist in a permanent abnormal dilation and damaging of the airways. Other

respiratory diseases can also lead to bronchiectasis, whether a severe lung infection during childhood, such as pneumonia, allergic responses associated with asthma, or in comorbidity with cystic fibrosis (characterised by production of thick mucus) and gastric reflux (British Lung Foundation, 2012).

Due to its symptoms of frequent sputum production, dyspnoea and wheezing, bronchiectasis can be easily linked with COPD or mistaken for COPD, which imposes extra urgency in the correct diagnosis of the disease and in choosing the most appropriate treatment.

Table 1.10 | Individuals diagnosed with bronchiectasis over the period of 2008-2012, in the UK

The highest incidence of bronchiectasis cases was registered in the West Midlands of the UK, as opposed to South East of England, which indicated the lowest prevalence of the disease (British Lung Foundation, 2012).



People suffering from bronchiectasis are more prone to bacterial infections due to the mucus accumulation in the bronchi, which consequently worsens the symptoms, sometimes resulting even in haemoptysis (coughing up blood), chest pain and inflammation of joints.

Individuals suspected of bronchiectasis, usually undergo a computerised tomography (CT) scan or CAT scan, in order to confirm the unusual widening of the airways and the diagnosis. Further cytological tests of a sputum sample may be required to identify the existence and type of bacteria present in the mucus and a blood test that may reveal the potential causes of bronchiectasis (British Lung Foundation, 2012).

1.7.3 | Treatment and Management of Bronchiectasis

Given the fact that bronchiectasis is an irreversible disease the available treatments aim to prevent additional damage and reduce the symptoms in order to improve the patients' quality of life. Bronchiectasis infections are treated with antibiotics, usually through oral administration in tablet form or inhaled through a nebuliser.

An alternative approach is pulmonary rehabilitation based on breathing exercises and methods that facilitate the reduction of sputum build-up in the lungs. If pulmonary rehabilitation does not meet its requirements, instead there are pharmacologic agents for mucus clearance. These consist of a hypertonic saline solution concentrated into nebulisers or Carbocisteine inhaler devices with the aim of maintaining the airways hydrated and facilitating the expectoration of mucus from the lungs (British lung Foundation, 2012).

Studies conducted on patients at a stable phase of bronchiectasis, have concluded that local and systemic inflammatory markers can indicate the state of the disease.

Additionally, evaluations of CT scans, lung function and health related quality of life (HRQoL) in bronchiectasis patients, have indicated increased levels of inflammation markers in their total white cell count, neutrophils and C-reactive protein levels. Further on, the elevated inflammatory markers were consistent with previous results from the CT scans, their lung function and the HRQoL (Wilson *et al.*, 1998).

1.8 | Pulmonary Sarcoidosis

Sarcoidosis as a general disease can affect any parts of the body including the skin, muscles, nervous system, nose, sinuses and eyes, the heart, lungs, liver or spleen. Although sarcoidosis has not been defined as a lung condition, in approximately 90% of cases the lungs or the tracheobronchial lymph glands are affected.

Pulmonary sarcoidosis is characterised by accumulation of granulomas (cell lumps) in the lungs, which in time affects the natural activity of the lungs and body (British Lung Foundation, 2012).

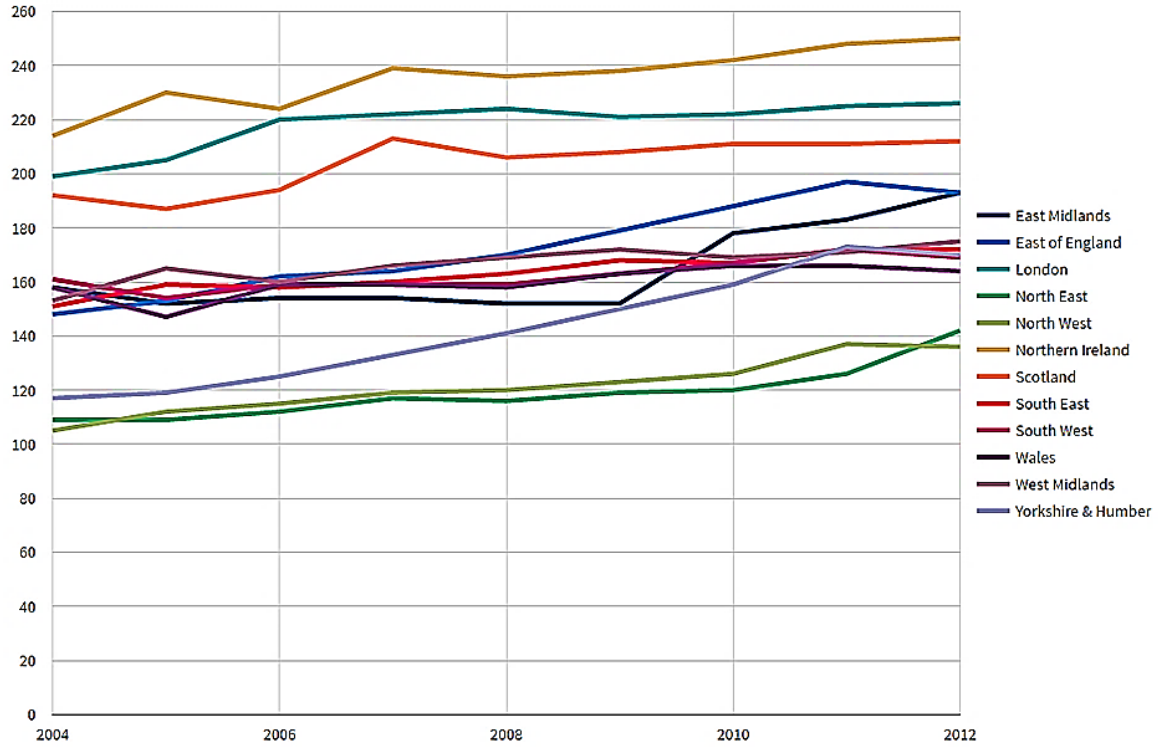
1.8.1 | Statistics in Pulmonary Sarcoidosis

Pulmonary sarcoidosis usually affects individuals with ages between 30 and 60 years old and is very rare in children. During 2008-2012 about 108,000 individuals in the UK developed pulmonary sarcoidosis. Based on the regional statistical prevalence of people in the UK with pulmonary sarcoidosis, between 2004 and 2012, the highest rates were in London, Northern Ireland and Scotland. These statistical results may be shaped by the influence of people of Irish

or Afro-Caribbean descent that predominantly live in London, Scotland and Ireland (Table 1.11).

Table 1.11 | Regional variation of people that have developed or lived with pulmonary sarcoidosis in the interval of 2008-2012

The prevailing cases of pulmonary sarcoidosis have been registered in Northern Ireland, London and Scotland. The outcomes may rely on the regional variation caused by Irish and people of Afro-Caribbean descent that predominantly live in Ireland, London or Scotland (British Lung Foundation, 2012).



Pulmonary sarcoidosis indicates similar incidences in both men and women, which rules out any differences based on gender. The mortality rates in pulmonary sarcoidosis indicate a total of 170 individuals that died in 2012, which accounts for 0.1% of all deaths from lung disease (British Lung Foundation, 2012).

1.8.2 | Causes of Pulmonary Sarcoidosis

The causes of pulmonary sarcoidosis remain unknown; however, the systemic nature of the disease has been indicated due to overproduction of white blood cells (Duchemann *et al*., 2016). It is likely that genetic susceptibility is a cause of sarcoidosis as it has been observed in monozygotic twins, the medical family history of patients that were currently suffering from the disease and the high incidence of the condition amongst specific ethnicities (Spagnolo and Schwartz, 2013). However, Spagnolo and Schwartz (2013) also argue that the pathophysiology of sarcoidosis may result in fact from a large variety of genetic and environmental factors operating simultaneously, each contributing with the precise proportions that can trigger the disease.

Depending on the disease severity, people suffering from acute pulmonary sarcoidosis may experience dry coughing, dyspnoea, enlarged tracheobronchial lymph glands, fever, hoarse voice or joint pains. Conversely, chronic pulmonary sarcoidosis develops gradually and presents fewer symptoms, but it can worsen in time (British Lung Foundation, 2012).

Giving the fact that pulmonary sarcoidosis shares some symptoms with other respiratory diseases such as bronchiectasis and IPF, diagnosis is dependent on a CT scan or chest X-ray or an endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA) procedure. EBUS-TBNA is a sensitive tool for the diagnosis of lung disease not visible with conventional bronchoscopy such as pulmonary sarcoidosis**.**

1.8.3 | Treatment and Management of Pulmonary Sarcoidosis

People with acute sarcoidosis usually do not require treatment, although for severe cases, steroids are used to improve breathing. Steroids are usually prescribed for a short period of time in high doses, followed by a lower dose over a long-term period. The steroid treatment

regime needs to be strictly followed otherwise the disease may regress and cause more damage. People with sarcoidosis are prone to accumulate increased levels of calcium, especially in their blood and urine. For this reason, calcium and vitamin D supplements are not recommended for individuals suffering from sarcoidosis as they should obtain these from a balanced diet (British Lung Foundation, 2012).

Duchemann *et al*. (2016) obtained metabolomic profiles of saliva from 24 patients suffering from sarcoidosis and 45 healthy controls via proton Nuclear Magnetic Resonance (NMR) spectroscopy. The results indicated several altered metabolites in the saliva samples of sarcoidosis patients in comparison to the healthy control samples, as well as a dysregulation of low levels of methanol and butyrate and high levels of acetate, lactate and N-butyrate. This study has confirmed that NMR metabolomics can distinguish specific biomarkers in saliva samples of sarcoidosis patients from the healthy individuals, which can be particularly beneficial in understanding the aetiology of this respiratory disease.

1.9 | Future Perspectives on Respiratory Diseases

The burden of respiratory diseases in the UK and worldwide, has proven to affect the economy, the healthcare system but most importantly peoples' lives. Regardless of the efforts that are being made in order to deal with this difficult task, mortality rates caused by lung diseases continue to increase year by year, leaving the effect of treatment and disease management powerless or with a transient effect.

A challenging problem for clinicians is determining the prognosis and diagnosis of respiratory diseases, due to their similarity in symptoms or co-morbidities.

Biomarkers from easily obtainable biofluids such as sputum, blood and urine have proven to be the key factor solutions that can be utilized to predict the clinical behaviour of pulmonary

diseases and render accurate diagnoses. Additionally, biomarkers offer the opportunity of observing the natural disease progression at molecular level, which in turn may provide a better understanding of its pathogenesis.

The utility biomarkers provide is the potential of determining the causes of disease progression and provide well-defined ways of disease classification for improved methods of prognosis and diagnosis in respiratory diseases.

The research described in this thesis will test if saliva is a reliable biofluid that can discriminate between respiratory diseases in comparison with controls through metabolomic and microbiomic analyses; and if it can represent a potential biosample for prospective studies with interest in obtaining biomarkers from it.

CHAPTER 2 | Biomarkers

The research can directly impact on clinical practice in the identification of biological markers (biomarkers) from biological fluids (biofluids).

Biomarkers have been utilised for diagnostic testing for more than 50 years and have established clinical value. Moreover, biomarkers offer the potential of assessing the treatment efficacy and aiding the development of new drugs, particularly in the area of therapeutic medicine for cancer (Gene-Quantification, 2013).

Formerly, biomarkers were defined as "cellular, biochemical or molecular alterations that are measurable in biological media such as human tissues, cells, or body fluids" (Gene-Quantification, 2013; Hulka, 1990). Nowadays the definition of biomarkers became more specific: "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention or other health care intervention" in accordance with the Biomarker Consortium of the Foundation for the National Institutes of Health (FNIH) in 2011. Thus, a given biomarker should be precise enough to reveal an exact biological characteristic or a measurable alteration in an organism, and is linked directly with a physiological condition or disease status. Biomarkers are beneficial practically for the entire course of a disease development. Before diagnosis, they offer the possibility for screening and risk assessment; during the diagnostic process, biomarkers are crucial for determining the stage and grade of the disease for the selection of the initial therapy; lastly, during the treatment period, they are helpful for monitoring the therapy outcome (Atkinson, 2001).

Biological markers from biofluids such as sputum, serum, plasma and urine provide early disease detection. This in turn offers an ideal opportunity for the improvement of patient

prognosis, streamlining drug therapy use, as well as evaluation of the clinical outcomes of specific treatments (Gene-Quantification, 2013).

Formerly, the biomarker identification has involved laborious investigations of defined cells, genes, proteins or biochemicals.

Recently the identification of biomarkers involves multi '-omics' layers or any other phenotypic measures regarding an individual's health status. The use of '-omics' (proteomics, metabolomics, microbiomics, genomics, etc.) and other high throughput technologies has allowed researchers to rapidly measure many variable and in this way, identify more than in the past decades. Therefore, these analytical improvements and high-level technologies using the '-omics' expertise, offer the advantage of generating more candidate biomarkers with possible clinical value. Coupled to this has been the development of bioinformatical tools which are integrate large data sets that derive statistically robust biomarkers linked to such as disease pathways and the pharmacologic outcome of medicines.

Biological markers aid in the understanding of disease progression, which gives them a crucial role in helping shape any medical treatment strategies. This in turn, may facilitate in tracking a disease from its initial manifestation until the terminal stage and conduct a variety of clinical technologies and functions such as molecular diagnostics, drug discovery, clinical trials and complex bioinformatical analysis (Gene-Quantification, 2013).

2.1 | Saliva

In the last century, saliva has emerged as a viable alternative to blood sampling by allowing, in certain cases, cost-effective and rapid diagnoses. Notably, saliva, exemplifies an ideal surrogate to blood for the identification of clinical biomarkers (Zhang *et al*., 2009). These can

be further used for clinical applications in monitoring the health status, the disease onset and progression of a patient.

Saliva is produced and secreted from the three-paired extrinsic salivary glands (Figure 2.1). The basic secretary units of salivary glands (acini) secrete a fluid that contains water, electrolytes (Wotman *et al.*, 1967; Dawes, 1969), mucus (Watanabe, 1948, Miglani, Raghupathy, 1968) and enzymes (Meites, Rogols, 1971; Pigman, Reid, 1952; Raus, Tarbet, Miklos, 1968). The fluid then flows out of the acinus into the collecting ducts, where its composition is altered. Finally, the small collecting ducts within the salivary glands join into larger ducts, which eventually form a single duct that drains into the oral cavity (Schneyer and Schneyer, 1964).

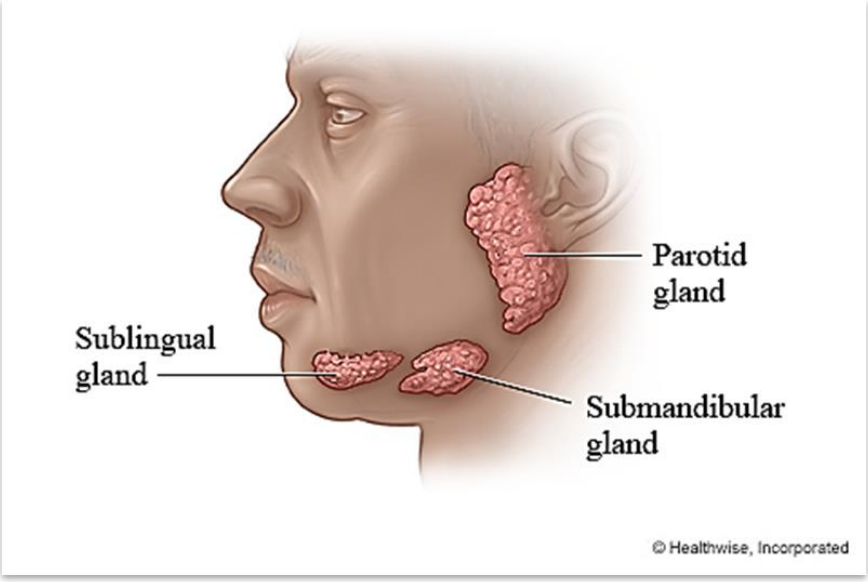


Figure 2.1 | Representation of the salivary glands
The two parotid glands represent the largest salivary glands, positioned all around the mandibular ramus (in humans) and produce 20% of the total salivary content in the oral cavity. The submandibular glands, located beneath the lower jaws, produce approximately 65-70% of saliva, despite the fact that they are much smaller than the parotid glands. The sublingual glands, situated inferior to the tongue, anterior to the submandibular glands and approximately 5% of saliva from these, enter the oral cavity.

Human saliva consists of 95-99.4% water and various minerals plus electrolytes (such as potassium, bicarbonate, sodium, calcium and chloride), mucus, white blood cells, epithelial cells, glycoproteins, enzymes and antimicrobial compounds such as Immunoglobulin A (IgA), lactoferrin and lysozymes (Farnaud *et al*., 2010).

Measurement of total organic constituents in salivary electrolytes, comprise sodium and chloride that contain a mmol/L concentration lower than plasma; calcium which has a mmol/L concentration similar to plasma; iodine, potassium and bicarbonate with a mmol/L concentration higher than plasma, and ultimately magnesium and phosphate (Boron, 2003).

2.2 | Validation of Salivary Biomarkers

Saliva is very attractive as a source of biomarkers as it can be obtained with minimal anxiety and discomfort to the patients and will thereby increase their willingness to undergo health inspections (Zhang *et al*., 2012; Wang *et al*., 2012; Kumada *et al*., 2007). The easiness and rapid collection of uninduced saliva offers the opportunity of obtaining real-time levels of biomarkers, which constitutes an advantage factor in comparison with other biofluids such as urine (Soo-Quee Koh and Choon-Huat Koh, 2007). However, previously, the use of human saliva as a diagnostic biofluid constituted a barrier, as it was believed that it possessed less informative analytes (or at least, in lower concentrations) than present in serum. However, Zhang *et al*. (2013), suggests that whole saliva should be used more frequently for diagnosing various human diseases because it is readily collected and even more, contains serum constituents as well. Interestingly, Bibby (1949) and Miletich (2010), highlight that since ancient times, doctors of traditional Chinese medicine have suggested that saliva and blood declaring that these biofluids come from the same origin.

The number of proof-of-principle assays based on saliva, has increased in the past years, which suggest this biofluid is relevant for the monitoring diseases and such as immune responses to viral infections or detection of illicit drug use (Zhang *et al*., 2013; Wu *et al*., 2011; Nagler *et al*., 2007). This has offered the possibility to discover a large number of medically valuable analytes, which assist in the identification of saliva biomarkers for different diseases including cancer, autoimmune diseases, viral and bacterial diseases, cardiovascular diseases and HIV (Epstein *et al*., 2000). In addition, approximatively 20% of the total salivary proteins were also identified in plasma, both fluids indicating disease-linkage and comparable functional diversity in their proteins (Bandhakavi *et al*., 2009).

A research study reviewed by Zhang *et al*. (2009) that compared saliva, blood and urine, concluded the fact that the diagnostic power of saliva can be comparable with that of blood, as it harbours approximately 30% of the same proteins found in blood. Similarly, Perez-Cornejo *et al*. (2012), assert that saliva is as beneficial for clinical applications as blood and urine and even more, it is easier to process it during laboratory procedures because it does not clot, like blood. Moreover, in their later study Zhang *et al*. (2013), confirms that saliva is indeed recognised as an effective diagnostic medium for detection of disease biomarkers through the application of proteomic analyses.

Following on from such work, salivary diagnostic approaches have only been recently developed and assessed to provide scientifically plausible, cost-effective, non- invasive and convenient detection methods for clinical applications. Thus, salivary biomarkers have proven to be extremely efficient in monitoring and predicting other medical conditions as cognitive impairment and Alzheimer's disease (Liang *et al*., 2016) or occupational stress (Soo-Quee Koh and Choon-Huat Koh, 2007). When examining the influence of stress in occupational health on individuals affected by anxiety, depression, exhaustion and staff turnover, Soo-Quee Koh

and Choon-Huat Koh (2007) have indicated a correlation between stress and increased levels of certain salivary biomarkers. The increased levels of salivary biomarkers caused by chronic or acute stress were cortisol, immunoglobulin A (IgA), lysozyme, salivary α-amylase and lastly, chromogranin A. Interestingly, smokers showed higher levels of stress biomarkers, compared to the non-smokers. Suzuki *et al*. (2016) had similar findings in a research based on the biomarker levels of salivary stress in cigarette smoking young adults. Results from profiled saliva samples indicated increased concentrations of cortisol, secretory immunoglobulin A (SIgA), interleukin-1β (IL-1β) and TNF-α in association with mood states of tension-anxiety, depression, hostility, fatigue and confusion. Saliva also holds a potential intrinsic value for detecting proliferation disorders, such as cancer (Coller, 2014). Moreover, Iorgulescu (2009) reveals that changes in saliva represent an indicative of the wellness of the patient.

2.3 Biomarkers in Saliva for Pulmonary Diseases

Many lung diseases can cause changes in certain chemicals in individuals' saliva, sputum, blood and urine. If such biomarkers are detectable sooner or easily, before a patient develops any specific symptoms it impacts on the treatment of pulmonary diseases (Lin *et al*., 2013).

At present, it is known that sputum samples originate from the central respiratory airways and its microbiological components (e.g. viruses, bacteria), cellular components (e.g. eosinophils, neutrophils) and protein components (e.g. mucins, cytokines) (Nicholas and Djukanovic, 2009). Hence, sputum is recognised as an efficient candidate for monitoring and assessing lung diseases.

There is a pressing need to discover non-invasive new techniques of detecting lung cancer in its early stage. For this reason, early detection represents one of the most feasible and promising approaches to moderate the increase of the lung disease (Xiao *et al*., 2012). Current progresses

in the field of salivary biomarkers, show valid outcomes designed for the analysis of lung cancer transcriptomes (Zhang *et al*., 2012a). Based on the logistic regression model that combined five messenger ribonucleic acid (mRNA) biomarkers, salivary transcriptomes of lung cancer patients were compared to the normal control patients. Results showed 93.75% sensitivity and 82.81% specificity in the pre-validation sample set with a calculation of the area under the curve (AUC) value of 0.93 (Zhang *et al*., 2012). This in fact, proves that the salivary mRNA biomarkers carry the potential for detecting lung cancer. Additionally, following Xiao *et al.* (2012), when profiling proteins in saliva, potential biomarkers were revealed, indicative of cancerous cells. Thus, the candidate biomarkers in lung cancer patients compared to the healthy control subjects, displayed in the pre-validation sample set 88.5% sensitivity and 92.3% specificity with AUC =0.90 (Xiao *et al*., 2012).

In consequence, these results suggest that a simple saliva test may be proficient enough for the screening and detection of various diseases, but most importantly it may be applied for serious lung diseases, such as lung cancer with the aim of detecting it early and increasing the five-year survival rate of individuals. This in turn, can be used extensively for future medical diagnostics.

Following the work of Lee and Wong (2009), saliva is recognised as containing microorganisms found in the mouth and sputum contains microorganisms found in the lungs. Such microbiological, cellular and protein components found in sputum can be used as markers of disease severity, exacerbation, susceptibility or progression. These can offer specific information regarding the patient's inflammatory cell content as well as physiological properties of sputum, such as purulence and bacterial load. Similar assessments have not been made of saliva samples.

The main challenge of this project is to validate the hypothesis that saliva can differentiate between pulmonary diseases through metabolomics and microbiomic assessments. Subsequently, when linked to other more in-depth research, the results of these studies will constitute in a large body of knowledge displaying characteristics in both healthy subjects and patients suffering from pulmonary diseases, based on saliva only.

CHAPTER 3 | Assessment of the Human Salivary Metabolome in Pulmonary Diseases

3.1 | Introduction

Metabolomics is one of the new "omic" sciences and a metabolomic study represents the evaluation of endogenous small-molecules called metabolites within a biological sample.

Metabolomics analyses have become a pivotal platform for their potential of discovering clinically relevant biomarkers within a biological system, that are simultaneously affected by a disease (Zhang *et al*., 2012b). Metabolomics utilizes two principal technological methods: mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. Both MS and NMR can be used for metabolomics fingerprinting and profiling, according to their extent of applicability, each with their own advantages and limitations (Madsen, Lundstedt and Trygg, 2010; Emwas, 2015). Additionally, NMR spectroscopy is a quantitative method, which does not necessitate the additional separation and derivatization steps for the sample preparation as in MS, although it is limited in sensitivity. Mass spectrometry, on the other hand, provides an excellent analysis approach for metabolomics research because it combines a sensitivity and selectivity platform, as well as different ionization techniques and mass analyser technology, which in turn increases the detectability of metabolites (Emwas, 2015).

The human metabolome comprises all of the metabolites in the body which are vital physiological and pathological indicators of the state of a biological system. Additionally, the metabolite profiles constitute unique chemical "fingerprints" that can provide insights with respect to a disease's cause, state and progression, through the identification and assessment of markers from biofluids (Zhang *et al.*, 2012b; Zhang, Sun, Wang, 2012).

The Human Metabolome Project is an attempt to unify metabolomics approaches that followed similar methodologies starting from the sample preparation and carried out until the final biological interpretations (Madsen, Lundstedt and Trygg, 2010). According to Zhang *et al.* (2012d), in order to acquire an accurate representation of the human metabolome or any other biological system, the use of two or more MS techniques may render better results. At the moment, there are no methodological approaches able to analyse the complexity of metabolites within the whole human metabolome. The mass spectrometry technological methods in metabolomics will indicate metabolites based on their mass-to-charge ratio (*m/z*) in charged particles of a biofluid (Spratlin, Serkova and Eckhardt, 2009).

A mass spectrometry technique, also employed in this study is the flow infusion electrospray mass spectrometry (FIE-MS). This method addresses several issues that are usually problematic when achieving adequate metabolome coverage due to the deficiency regarding the instrument dynamic range, mass resolution, degree of sensitivity and signal identification (Draper *et al.*, 2013). In the context of FIE-MS, sputum metabolite changes with an AUC greater than 0.8 were identified only in lung cancer patients when compared to healthy controls (Cameron *et al.*, 2016). A recent study in the field (O'Shea *et al.*, 2016), based on profiled sputum samples from patients with NSCLC and SCLC and age-matched smoking controls utilizing FIE-MS, has indicated an 80% sensitivity and 100% specificity for predicting SCLC. Furthermore, the metabolomics profiling identified six candidate biomarkers, which have already been associated with lung cancer. For instance, the enzyme glycine decarboxylase (GLDC), which is responsible for the degradation of glycine has been linked to NSCLC due to increased levels of GLDC and very low degree of glycine found in individuals suffering from non-small cell lung carcinoma (Zhang *et al.*, 2012e).

These examples alone highlight the potential of FIE-MS metabolic profiling of biofluids and the superior results they deliver, especially for the scarcely exploited paths of saliva biomarkers

for lung cancer. Additionally, flow infusion electrospray mass spectrometry may be an essential contributor in the whole process of metabolite detection for improved clinical diagnostic pathways in patients with lung cancer and other pulmonary conditions.

Although the metabolomics techniques of analysis generally produce considerable amounts of data, for experiments focused on understanding the disease dynamics within a human metabolome, only certain parts of the whole metabolome may exhibit alterations. A study based on MS that involved the profiled saliva samples from 215 patients suffering of breast, oral and pancreatic cancers, managed to identify no more than 57 metabolites related to specific alterations linked to either of the three cancers (Sugimoto *et al.*, 2010).

Despite some significant advancements that have been made with the contribution of metabolomics in the field of cancer, unfortunately these cover only a few that focus on the lung cancer. However, this is not the case for all respiratory diseases to impose the same difficulty, since COPD is being particularly studied. Indeed, metabolites changes that are representative for all COPD GOLD severity stages have been defined (Ubhi *et al*., 2012).

3.1.1 | Objectives

Generally, saliva is one of the least studied biofluid in lung cancer metabolomics. However, the potential of saliva as a biomarker for other respiratory diseases has started to receive more attention in the past decades. This is due to the complexity of its constituents' non-invasiveness method of collection, and most importantly direct link to the respiratory tract.

Provided that saliva depicts specific symptoms of the inflammatory lung airways, this can be further employed in metabolomics profiling techniques. This may offer better insights regarding the role of salivary metabolites in the aetiology, progression and treatment of respiratory diseases.

The main aim presented in this segment of study was to explore the salivary metabolome from eight patients with confirmed lung cancer and 47 COPD patients contrasted with 37 “healthy” controls. This was done in order to demonstrate that saliva as a biofluid, allows discrimination between pulmonary diseases when employed in metabolomics analyses. To this, 16 patients suffering of asthma, three from pneumonia, one of bronchiectasis, two of IPF and three of sarcoidosis were added to the analysis to increase the database and add a greater variety to it. Moreover, the study investigations of all 117 processed saliva samples were carried out with the aid of the MetaboAnalyst 3.0 (Xia *et al*., 2015) analysis platform on diverse grouping variables such as diagnosis, disease severity and stage (i.e. GOLD stage), age group, gender, hospitals, smoking status, medical history and other symptoms.

3.2 | Materials and Methods

3.2.1 | Patient Recruitment and Saliva Sampling

This study was carried in conjunction with the Hywel Dda University Health Board with the ethical approval (16/WA/0036) from the loco-regional ethical committee and can be also accessed on the International Standard Randomised Controlled Trials Number (ISRCTN) registry under the study number ISRCTN16657101. All samples and data were anonymised prior to being processed and analysed. Moreover, this clinical study of saliva constitutes a division from the project “Novel Technologies for Detecting and Monitoring Pulmonary Diseases” based on other biofluids as sputum, blood (plasma and serum) and urine.

Patient recruitment took place between the time intervals of February 2016 to August 2016 during respiratory clinic attendance or recruiting from the wards at three hospital sites. All

uninduced saliva samples employed in this research study were collected from patients suffering from pulmonary diseases from three different hospital sites around Wales: Bronglais Hospital (BGH) from Aberystwyth, Glangwili Hospital (GGH) from Carmarthen and Prince Philip Hospital (PPH) from Llanelli with the help of research nurses and consultant doctors.

The first crucial step regarding patient recruitment, involved obtaining the patients' signed informed consent in duplicate so that one copy can be given to the consented person and the other to be attached in the patient's medical history file. The recruitment protocol consisted in a Participant Information Sheet (Appendix Chapter 3, Supplementary Figure 3.1), which was self-explanatory and contained information and frequently asked questions about the study, for the patient's better acknowledgement. Following, the Clinical Information Datasheet, which was designed as a questionnaire format with the aim of recording as much as possible information regarding the patient's medical status and diagnosis, age, gender, specific symptoms, disease severity (FEV_1/FVC for GOLD staging in COPD), medication/antimicrobial medication and past medical history/family medical history.

Every piece of information that was recorded constituted an important variable in the analysis and interpretation processes. Finally, the samples of saliva collected were transported and stored accordingly at -80°C in the research laboratories at the Institute of Biological, Environmental and Rural Sciences (IBERS) of Aberystwyth University.

Before analysing the data, all saliva samples have been classified into four groups based on patient diagnosis: LC, COPD (comprising emphysema and chronic bronchitis), "Other" category (including asthma, pneumonia, idiopathic pulmonary fibrosis, bronchiectasis and sarcoidosis), and the Healthy Control (HC) group (mainly composed of ex-, non- and current-smokers).

3.2.3 | Processing of Raw Saliva

The protocol was adapted from that of Cameron *et al*. (2016). In order to wash the glass beads (Sigma-Aldrich, Dorset, UK), 200 mg of glass beads powder were washed in 2 ml acetone and placed on a vortex mixer (Vortex-Genie 2 Shaker) at the full speed of 2700 rpm for 30 seconds. Further on, the microcentrifuge tubes with the weighted 200 mg glass beads were centrifugated (Heraeus Multifuge 3SR) at 1800 x g for one minute and left uncapped in the fume cupboard overnight to dry out completely.

For the sample preparation, saliva samples were thawed on ice at 4˚C and 50 µl of the supernatant were individually transferred to a sterile 2 ml microcentrifuge tube to which 7.5 mg of ≤160 µM (particle size) acetone-washed glass beads were added.

To these, 380 µl of a solvent mix of high performance liquid chromatography (HPLC) grade methanol (MeOH) and chloroform ($CHCl_3$), in a ratio of 4:1 were added in order to dissolve the solute (saliva) and also to separate and isolate the polar from non-polar metabolites in the biological samples and stop the intracellular metabolism. For a better accuracy and consistency, the HPLC mix was prepared in a large quantity of approximatively 100 ml and stored at room temperature in a Duran bottle, so that all 117 saliva samples were treated with the same solvent. The HPLC grade mix represents a helpful step in metabolic fingerprinting, as it acts as an extracting agent by making the cell envelope permeable and thus, triggering the release of intracellular metabolites from the saliva samples (The Metabolomics Lab, 2014).

To homogenise the mixture and disrupt the cells, saliva samples were placed on a vortex mixer for five seconds. Following this, samples were shaken for 20 minutes at room temperature and then stored at -20˚C for 20 minutes to allow precipitation. Samples then underwent centrifugation (Thermo Scientific Heraeus Pico 21 Microcentrifuge) at 11,000 x g for six

minutes at 4˚C. Following this the supernatant was transferred to a sterile 2 ml microcentrifuge tube.

From this, 100 µl were transferred to a MS vial and capped. For a better accuracy and quality control validation, a vial with 100 µl of the gene pool (composed of a mixture of 1 µl of each sample used in the experiment) and a vial only with the HPLC mix were added to the analysis of the MS instrument too. In order to preserve the saliva samples and prevent bacterial growth, samples were stored at -20˚C until run, in a randomised order using an auto-sampler.

The following final steps were performed with the help of Dr Manfred Beckham – Aberystwyth University. Subsequently, for each sample, 20 µl were injected into a flow volume of 60 µl per minute water-methanol, in a ratio of 70% water and 30% methanol, using a Surveyor flow infusion electrospray mass spectrometry for high throughput, non-targeted metabolite fingerprinting (FIE-MS) (Thermo Scientific, MA, USA). Data acquisition for each saliva sample was done by alternating the positive and negative ionisation modes, throughout four different scan ranges (15-110 m/z, 100-220 m/z, 210-510 m/z, 500-1200 m/z) (Draper *et al*., 2013; Cameron, 2015), on an LTQ linear ion trap with an acquisition time of five minutes (Thermo Electron Corporation, CA, USA).

3.2.4 | Data Analysis

Following, raw data outputs from the MS instrument (FIE-MS fingerprinting) were converted into an open file format. Data normalisation consisted of Total Ion Current (TIC) normalisation, according to the following equation: TIC normalised data =[data/ sum(data) ÷ median(data)] as shown in Figure 3.1 (a, b and c).

Then, centroided intensity peaks were derived from their base64 elements for all m/z values above 55.0. Data was binned below 0.25 m/z values and mean centred. These were then sorted

from lowest to highest and referenced back to their relevant identities and labels using the metadata spreadsheet.

Finally, data was base-10 log transformed and the normalised abundances were subsequently analysed using the MetaboAnalyst 3.0 platform (Xia *et al*., 2015) as displayed in Figure 3.1 (d, e and f). Following, the unpaired samples with their corresponding diagnoses were arranged in rows and the features were arranged in columns [peaks (mz/retention time(rt))]. The uploaded file was in comma separated values (.csv) format with numeric values only exported from the Microsoft© Excel™ 2013 spreadsheet.

In order to minimise variance and preserve the integrity of metabolites, the key elements for data integrity check, outlier detection, quality control, normalisation and scaling, were kept as similar as possible for each completed analysis. The purpose of this approach was to make the technical replicate groups (saliva samples) comparable with each other, yet without introducing too much variation between them. The profiling method used was the non-targeted chemometric one.

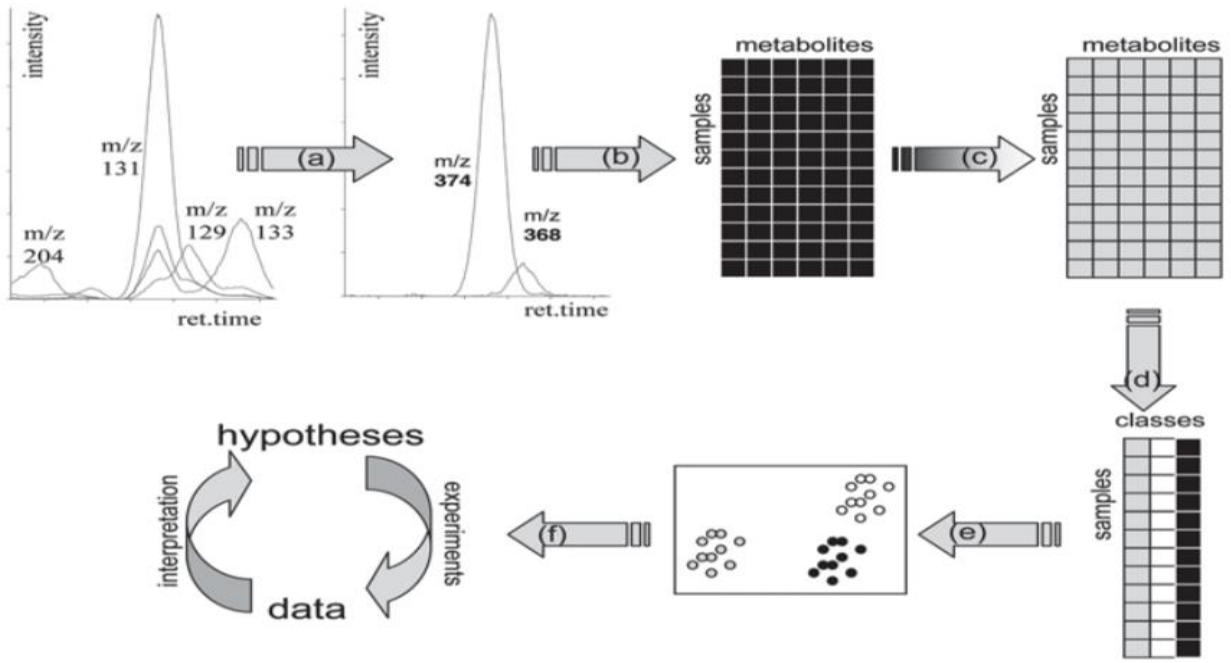


Figure 3.1 | Data analysis flowchart following flow infusion electrospray mass spectrometric analysis of saliva profiles
(A) Select optimal quantifier ions by centroid intensity peaks derived from their base64 elements for all m/z values above 55.0. **(B)** Sort peak areas lowest to highest and export these back to their relevant identities and labels in order to receive a data matrix (metadata). Data was binned below 0.25 and mean centred. **(C)** Normalise and transform data by Total Ion Current (TIC): TIC normalised data =[data/ sum (data) ÷ median (data)]. **(D)** Structure data according to background information regarding each sample and diagnosis class. **(E)** Obtain significant and unknown data structures by univariate and multivariate statistics. **(F)** Interpret findings, verify the hypotheses or generate new ones.

For the data filtering, which aims spectral alignment or binning, the interquartile range (IQR) analysis was chosen. The purpose of the IQR is to remove the outliers (highest and lowest numbers in a set) by selecting the mid-spread (middle 50) of a set of numbers.

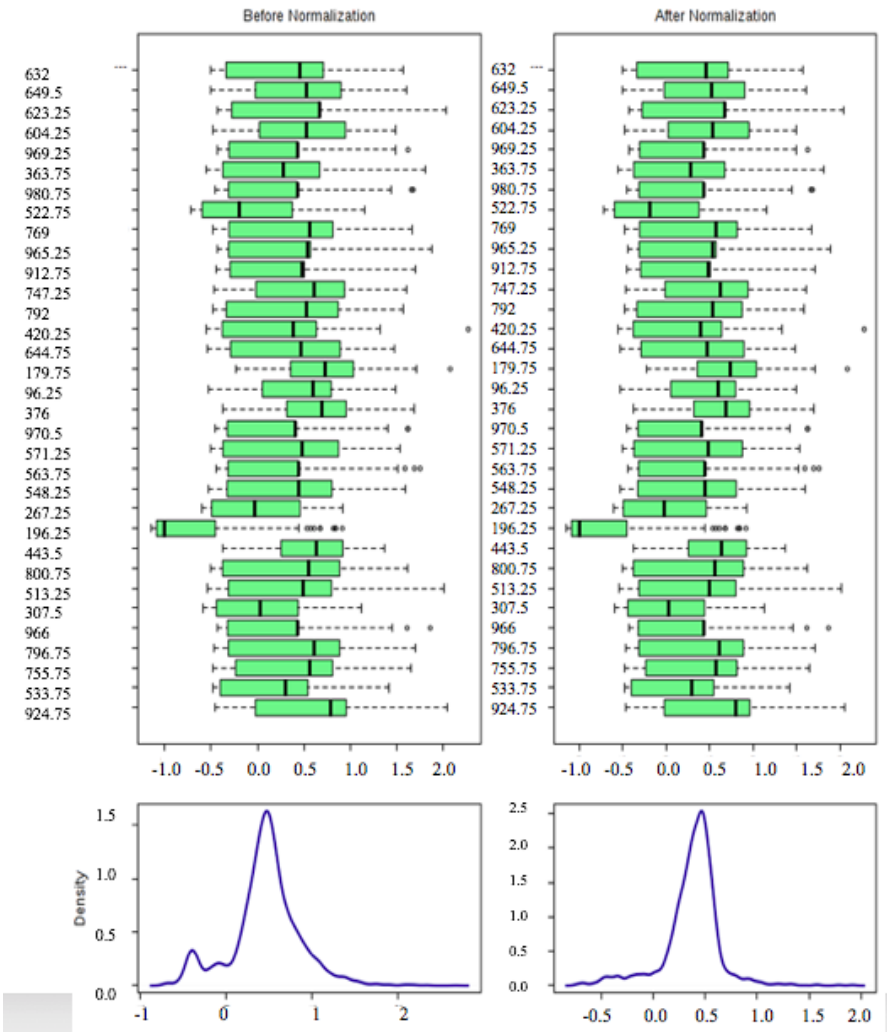


Figure 3.2 | Data Normalization of the COPD, LC, Other and HC classes
Normalisation of data comprised of four groups containing COPD, LC, Other and HC using the interquartile range (IQR) data filter. Sample normalization was employed for all analyses, as shown here for the first investigation of the four classes including LC, COPD (comprising emphysema and chronic bronchitis), Other (asthma, pneumonia, IPF, bronchiectasis and sarcoidosis) and HC.

Subsequently, data normalisation, which allows general–purpose adjustment for differences among samples was applied to all analyses to convert them to a normal distribution. The rationale of choosing these settings was to preserve the metadata in its initial format as much as possible, thus introduce less systematic variation and improve the performance of downstream statistical analysis, so the outcome would be as accurate as possible, displayed in Figure 3.2 (Katajamaa and Oresic, 2005).

3.3 | Results

3.3.1 | Metabolomics Analyses of LC, COPD, Other and HC

The initial analysis performed with the MetaboAnalyst 3.0 platform was conducted on a data matrix of 117 samples each with 3771 peaks (mz/rt). These comprised four groups: LC, COPD, "Other" group including other pulmonary diseases e.g.: asthma, sarcoidosis, idiopathic pulmonary fibrosis (IPF) and bronchiectasis, and healthy controls (HCs) including never-, ex-, non- or current-smokers.

The initial assessments tested whether there were any differences between the three classes of respiratory diseases (COPD, LC and Other) in comparison with the HC. In this way, any significant patterns could be determined with the use of one-way analysis of variance (ANOVA) or t-tests and chemometric multivariate analyses as the principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) or heatmaps.

Supplementary information regarding participant diagnosis, age group, smoking status and FEV1% predicted, are detailed in Chapter 3 Appendix, Supplementary Table 3.1, Table 3.2, Table 3.3 and Table 3.4.

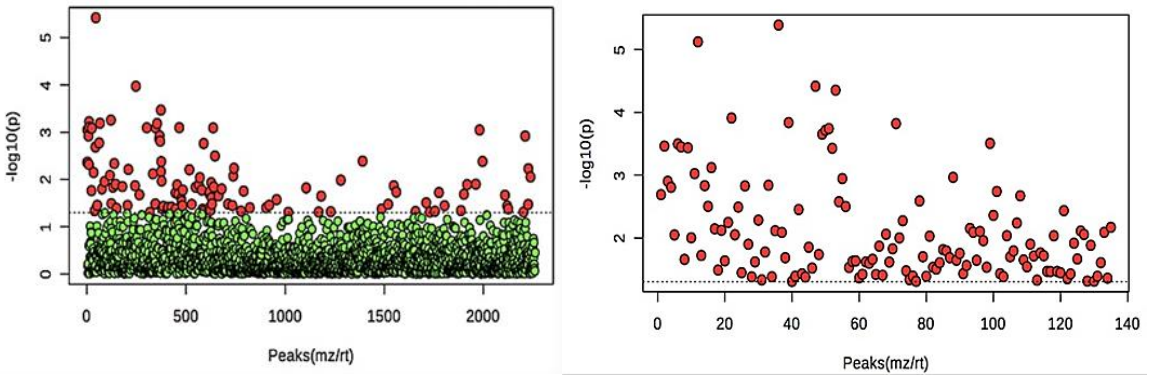


Figure 3.3 | ANOVA analysis of saliva samples from four different groups COPD, LC Other and HC

The post-hoc significance comparison column from the left-hand side, provided a preliminary overview regarding features that were automatically selected with the cut-off threshold, approximately higher than 1.4. The right-hand side however, displayed only the important features of the three classes of COPD, LC and HC significant in discriminating between the diagnoses under study that had a P value threshold of 0.05.

As shown in Figure 3.3 the ANOVA test differentiated between the samples analysed with the cut-off threshold scores of approximately higher than 1.4 and the significance plot of the P value threshold 0.05.

The second ANOVA test was performed with a new matrix that excluded the Other diseases class of samples. This was done as this represented a highly variable group. In this case, the analysis was focused more on the classes with patients that had only one diagnosis per group (COPD, LC and HC) and showed significance above the threshold.

Principal Component Analysis was employed to visualise difference between the groups (Figure 3.4); one which included the "Other" category and another where this was excluded.

Only when the "Other" category was excluded was partial separation of the COPD, LC and HC evident. Interestingly, two outliers, which are indicated by the blue arrows in Figure 3.4, have been plotted out of their clusters in the PCA carried only on the three classes (COPD, LC, HC), where the same observed in the first PCA conducted on the four classes (COPD, LC,

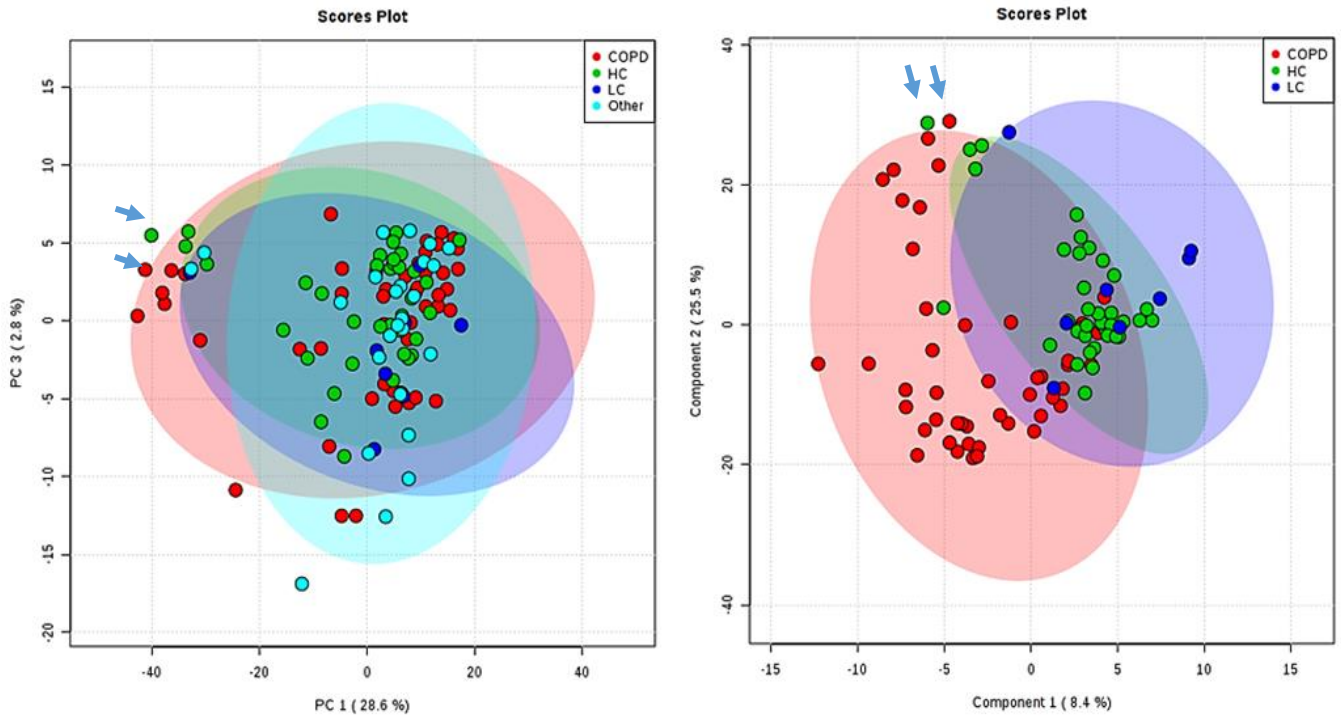


Figure 3.4 | Principal component analyses of COPD, LC, Other and HC
Principal component analysis between the matrix containing all four classes of diagnoses LC, COPD, Other and HC and the matrix composed of the significant values post ANOVA test, containing the COPD, LC and HC classes only. As expected, by removing the Other diagnoses classes and using only the significant samples selected by the ANOVA test, the second PCA (Component 2) seems to discriminate better between the three classes of COPD, LC and HC. The outliers (indicated by blue arrows) have remained constant in both analyses.

The HC outlier is represented by sample A122, which was noted as a female current smoker belonging to the 40s-age group recruited during a General Respiratory Clinic. The outlier of this sample was closely positioned to an A067 COPD saliva sample, represented by a female ex-smoker belonging to the 50s age group with a value of 28 pack years and suspected infection. One reason that may have caused these two saliva samples to be plotted out of the cluster might have been influenced by common variables as smoking and medical history including hypertension (HTN) and probability of a respiratory infection. Furthermore, an unusual overlapping is displayed between the LC and HC groups, and partially with the COPD group. Otherwise, the PCA carried on the three classes showed there is evident metabolite

variation between COPD patients in comparison with LC and HC patients. The LC and HC classes overlap (as displayed below in the following analyses as well) due to a possible similarity in metabolites.

Partial least squares-discriminant analysis was performed with the original PCA summarized variables (Figure 3.5). This produced better separation that in Figure 3.4. Similarly to the PCA, sample A122 represented an outlier once again as it correlated more with the COPD class. A new sample that constituted an outlier was a HC identified as A135 ex-smoker male belonging to the 80s age group with a total of 15 pack years and showing common symptoms of dyspnoea (highlighted by the blue arrows).

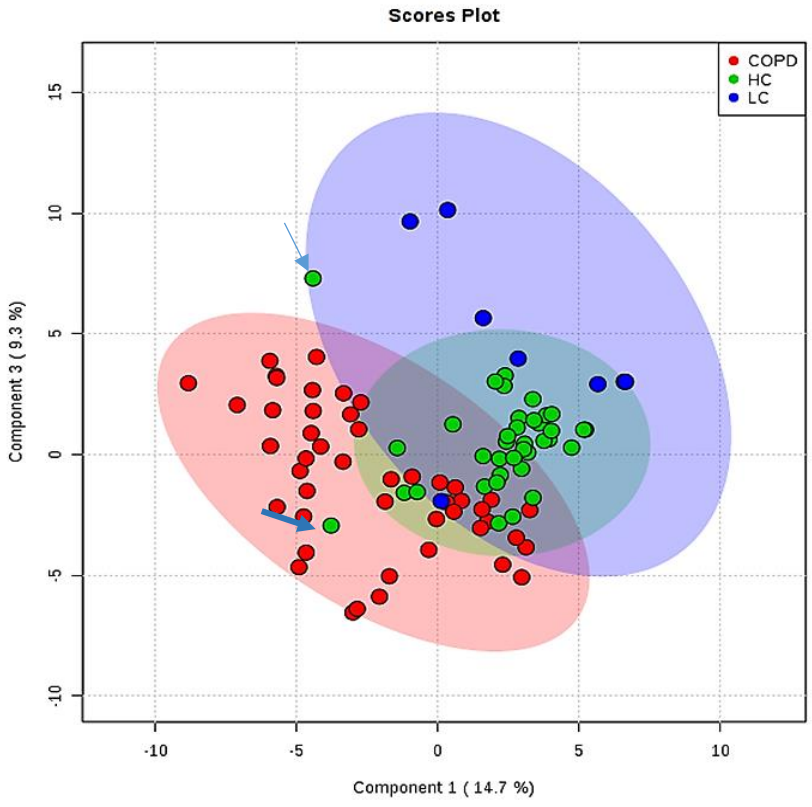


Figure 3.5 | PLS-DA performed on the three classes of COPD, LC and HC
The HC outlier (indicated by blue arrow) from the HC group presented a smoking history of 28 and 15 pack years and symptoms indicative of respiratory issues.

Subsequently, a Random Forest Plot (Figure 3.6) was used to represent the mean decrease accuracy values (VIP scores) of relevant metabolites identified in the four classes.

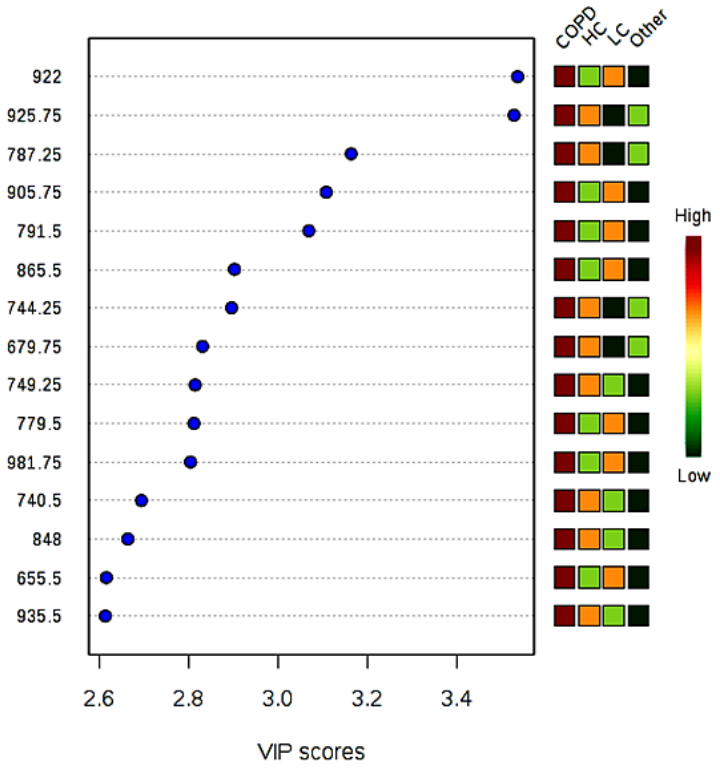


Figure 3.6 | Random Forest Plot displaying the VIP scores of relevant metabolites identified by MetaboAnalyst 3.0 in the four classes of COPD, HC, LC and Other
Detected values represent metabolites which have potential in terms of diagnostic markers and inflammation. As shown in the colour key: red signifies high or upregulated, yellow – medium and green shows low or downregulated, according to the VIP scores of a certain metabolite (vertical values on the left hand-side). Therefore, all identified metabolites here are upregulated in COPD samples, medium to downregulated in HC and LC, and downregulated in Other class.

Metabolites that are either higher or lower in all four classes are of main interest, because they reveal biochemical thresholds that can aid in understanding on what occurs at a molecular level during inflammation. This in turn can offer a greater insight with regards to exacerbation crises in COPD patients. As displayed in Figure 3.6, all metabolites were elevated in the COPD class, suggesting a high risk of exacerbations as it was also confirmed in the case of several patients. Healthy Control and LC patient samples displayed a combination of medium to lower levels in

metabolites, however the Other group has shown all decreased levels. A more in depth study should be allocated to the Other class with more data, in order to justify the low levels of these metabolites.

For a better perception of how saliva samples from patients with COPD, LC and Other diagnoses are related between each other and in comparison with the healthy individuals, every class of diagnoses has been analysed in tandem.

3.3.2 | Metabolomics Analysis of COPD and HC

Data analysis was carried exactly as in the previous subsection 3.3.1. The difference in the type of analysis was given by the fact that the group was restricted to two classes only, therefore, instead of ANOVA, a t-test analysis was performed. Fifty relevant variables were selected after the threshold cut-off and the P-values were transformed by –log10 so that more significant variables could be plotted at a higher scale on the graph (Figure 3.7).

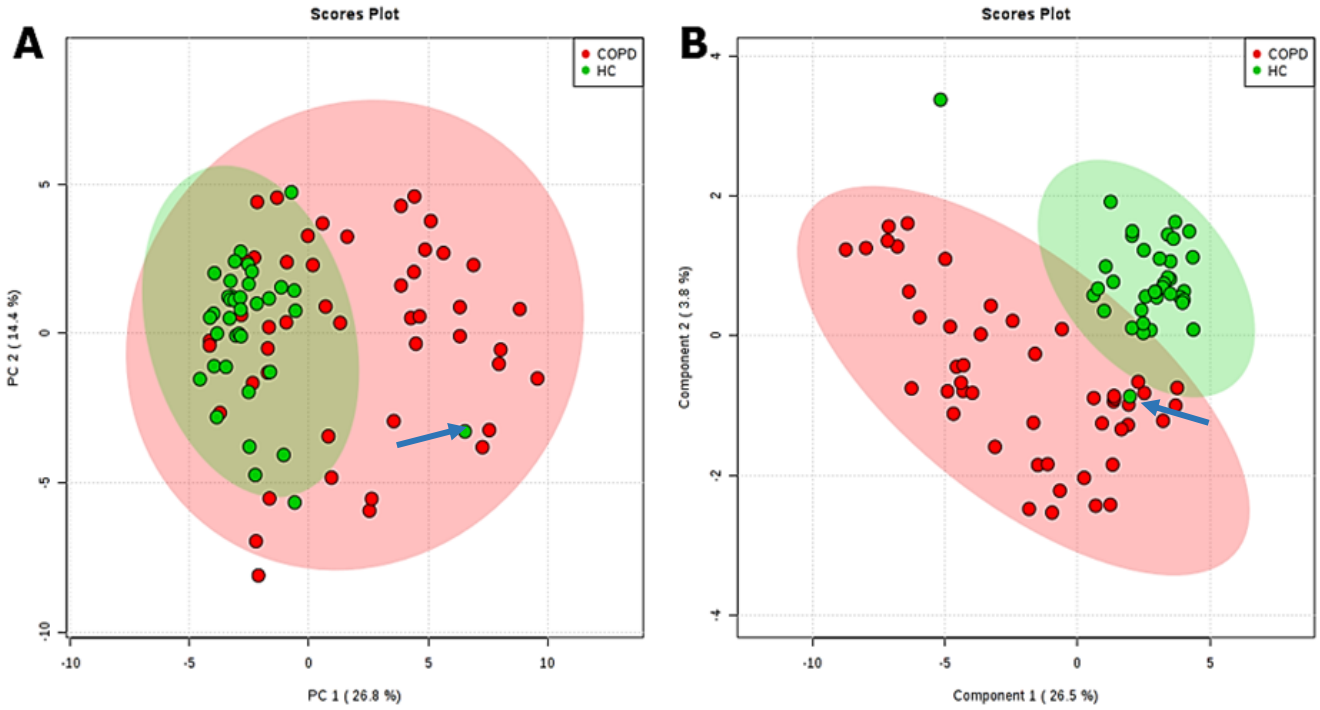


Figure 3.7 | PCA (A) and PLS-DA (B) of the COPD and HC classes with selected significant variables indicated after the t-test analysis
PCA and PLS-DA conducted on the HC and COPD classes revealed two outliers (indicated by blue arrows) from each group, which seem to show a repetitive separation when comparted to previous analyses.

The PCA has indicated one outlier which did not seem to be integrated in the cluster. This was represented by sample A067 from the COPD class, as in the previous analysis from subsection 3.3.1. Another interesting outlier was sample A135 from the HC class which seemed to indicate a trend as it was always in the COPD cluster. The PLS-DA showed a better separation of the COPD and HC classes, although an outlier has been identified in this analysis as well. Similar to the previous cases, sample A135 from the HC class was an outlier (indicated by blue arrows).

3.3.3 | Metabolomics Analysis of COPD and LC

The PCA and PLS-DA analyses between the COPD and LC classes (see previous subsection 3.3.1 for data analysis) were similar in partial cluster distribution, even for the outlier sample of the COPD class A067 as shown below in Figure 3.8.

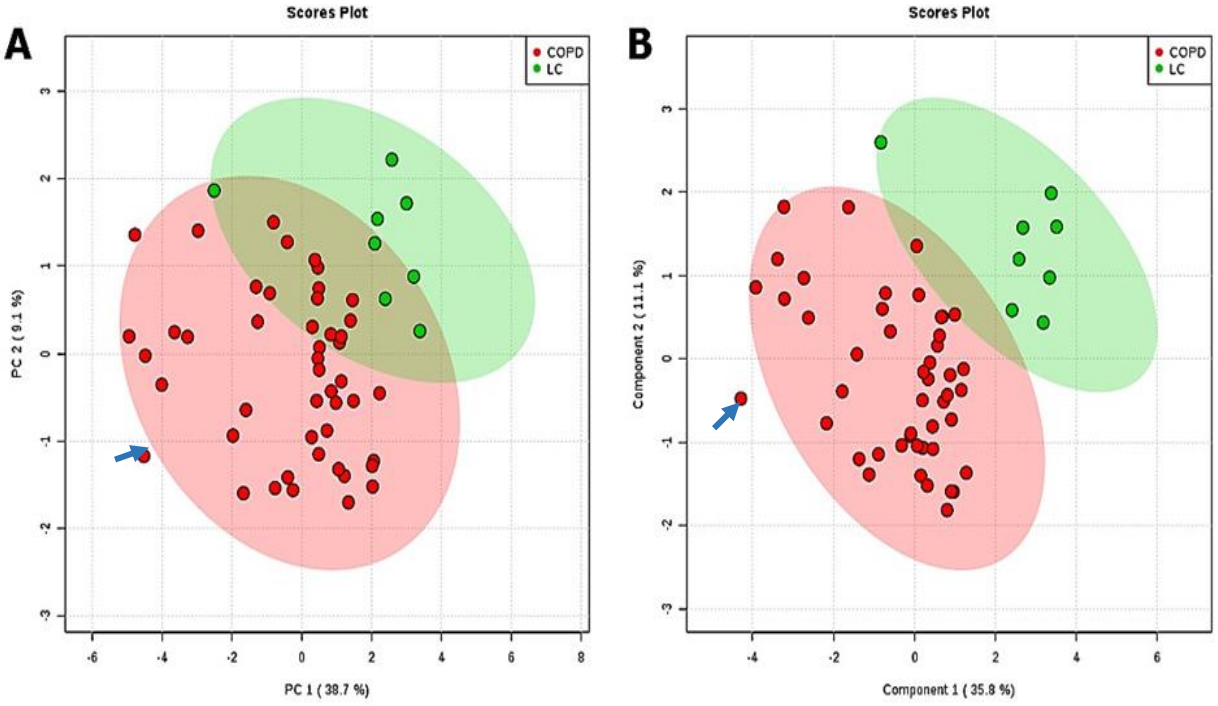


Figure 3.8 | Principal Component Analysis and PLS-DA between COPD and LC
Analyses between the COPD and LC classes performed with the significant values indicated by the t-test. PCA analysis is indicated in the A figure and the PLS-DA is indicated in the B figure. Both analyses have been distributed in concordance with Component 2 of the graph. Outliers are indicated by blue arrows.

3.3.4 | Metabolomics Analysis of COPD and Other

The analyses performed between the COPD and the Other diagnoses groups indicated similarity in both the PCA and PLS-DA tests (for data analysis see subsection 3.3.1) with only partial separation present. The PCA has shown an outlier from the COPD class represented by sample C017 male ex-smoker belonging to the 70s age group. This patient presented himself to the Glangwili Hospital with an exacerbation and this fact may play an

important role on the quality of saliva and thus, influence the value of the sample and further, the variable as well (Figure 3.9).

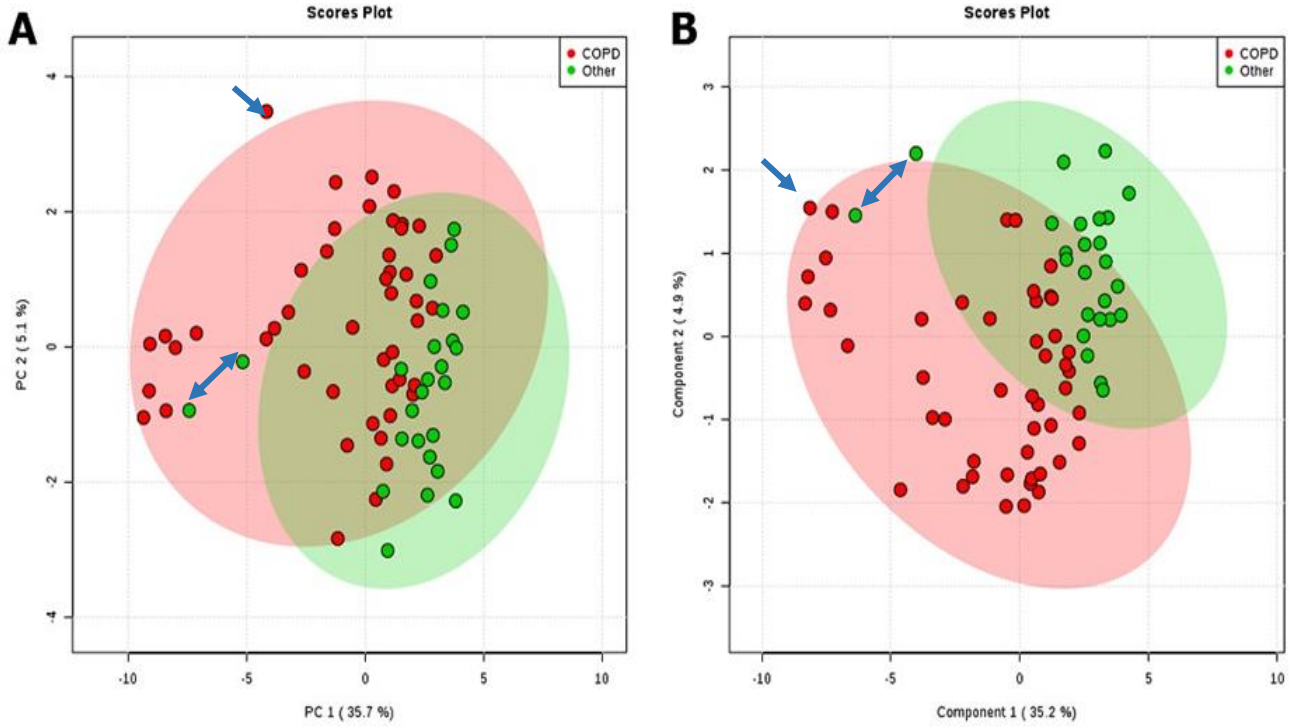


Figure 3.9 | Scores plots of PCA (A) and PLS-DA (B) tests performed between the COPD and Other diagnoses groups
Both types of analyses presented similar clustering with outliers represented by samples from patients that were suffering from exacerbations (indicated by arrows).

Moreover, both the PCA and PLS-DA recorded two new candidate samples as outliers which belonged to the Other diagnoses group and interfered in the COPD class. Sample A111 was represented by a non-smoker male from the 40s age group that was suffering of asthma.

The second outlier was represented by sample A108 which was an ex-smoker male belonging to the 60s age group which was suffering of asthma (often asthma exacerbations) and FEV1% of 54 (GOLD 2 or Moderate stage) and a total of 45 pack years.

These results suggest that outliers are indeed represented by the quality of the saliva samples, which in turn can reveal some similarity in symptoms of acute stages of respiratory diseases as COPD and asthma (Figure 3.9).

3.3.5 | Metabolomics Analysis of LC and HC

Data analyses of the matrix were done as described in subsection 3.3.1. The PCA test indicated less clustering between the two classes of HC and LC compared to the supervised PLS-DA. Additionally, the PCA result may have been influenced by certain outliers represented by the COPD class. Also, the separation between clusters on the PCA test, occurred mostly along Component 1 (Figure 3.10).

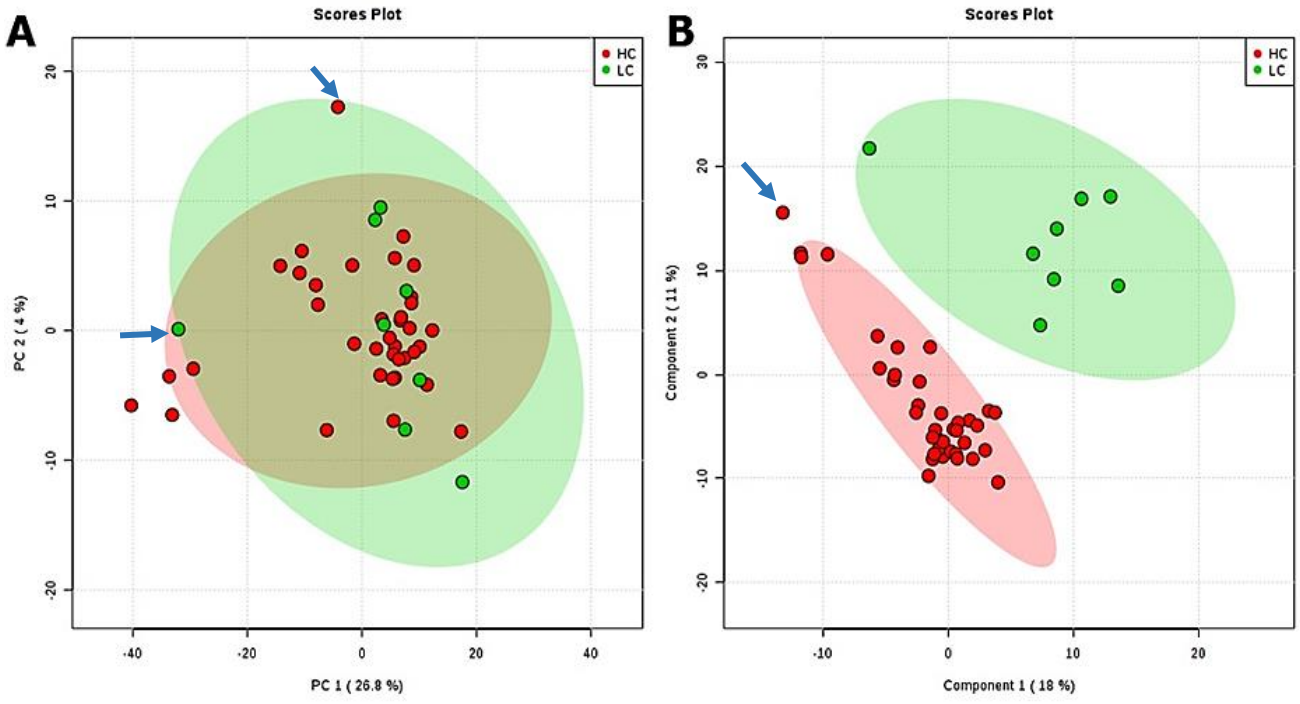


Figure 3.10 | PCA (A) and PLS-DA (B) tests between the HC and LC classes
Separation between clusters in the PCA test occurred mostly along the Component 1 with a few outliers, indicated by blue arrows, from the HC class.

Sample A122 discussed previously in subsection 3.3.1 continued to be a trend as an outlier of the HC class.

Another interesting outlier was represented by the HC sample A059 which was part of the LC cluster in the PCA test. Even if it was a HC, sample A059 was represented by a male current smoker with an 80 pack years and suspected respiratory infection.

The outlier from the PLS-DA test was again sample A122 as identified in the other previous analyses.

3.3.6 | Metabolomics Analysis of LC and Other

The PCA test showed an excellent separation between the LC and Other classes. Therefore, the supervised technique of the PLS-DA test was almost unnecessary (data analysis performed as mentioned in subsection 3.3.1). Both PCA and PLS-DA tests indicated obvious differences

between LC and Other classes. One of the detected outliers was the Other samples A111 discussed previously in subsection 3.3.4. The second outlier belonged to the Other class as well, and it was represented by sample C006 which was a female ex-smoker belonging of the 80s age group suffering of pneumonia (Figure 3.11).

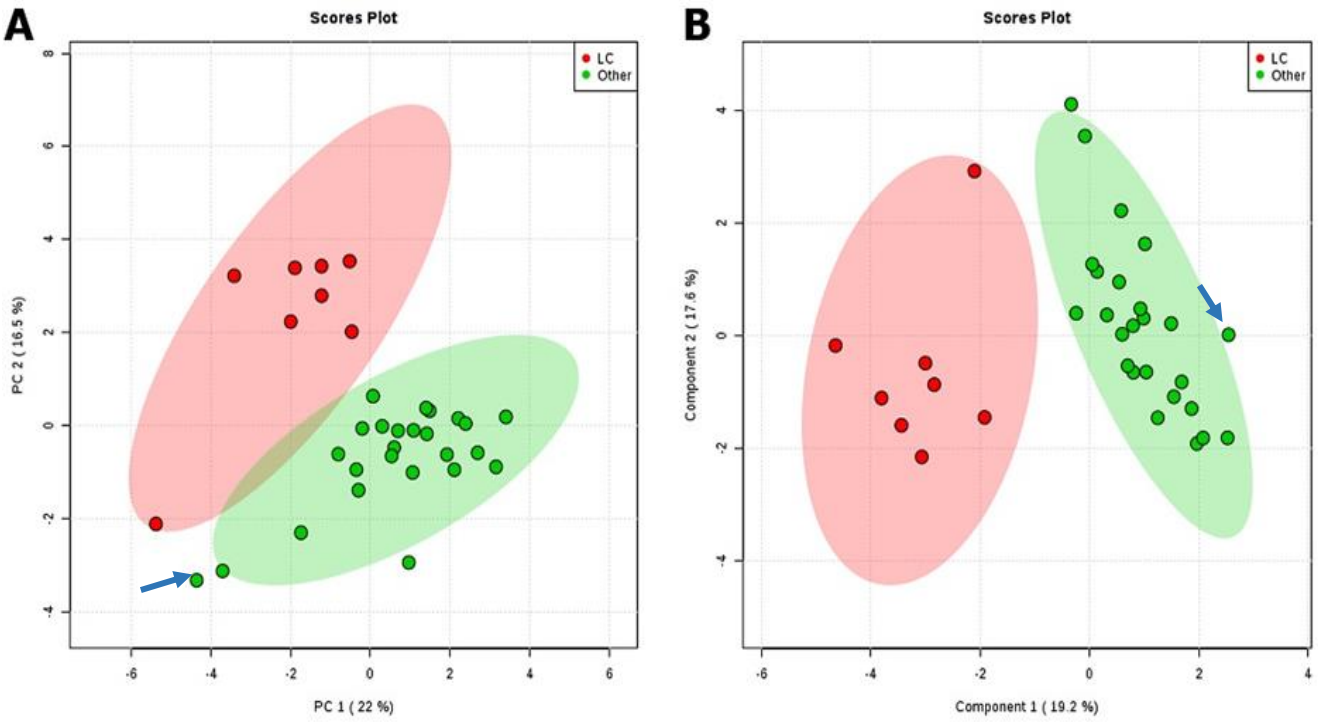


Figure 3.11 | Tests of PCA (A) and PLS-DA (B) between the LC and Other classes
Both analyses show an evident cluster separation with a few outliers belonging to the Other group. The two outliers (indicated by arrows) from the Other class include patients with a history of heavy smoking and co-morbidities. This can represent a valid motive with respect to their separation.

3.3.7 | Metabolomics Analysis of HC and Other

Data analysis was performed as described in subsection 3.3.1. Both the unsupervised PCA and supervised PLS-DA tests displayed similar clustering with partial separation and a few outliers represented by the HC samples A135 and A122 and the Other group samples A111 and A108 previously discussed in subsections 3.3.2, 3.3.1 and 3.3.4. Results are shown below in Figure 3.12.

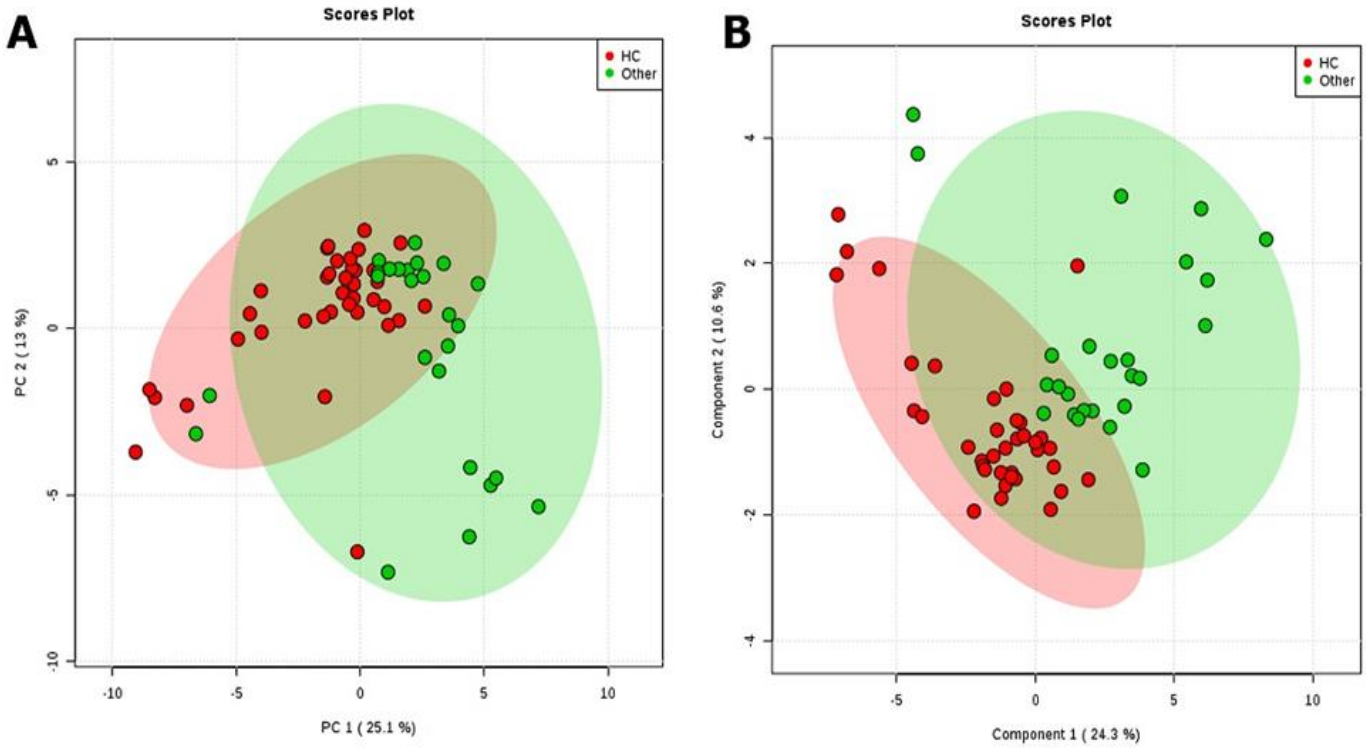


Figure 3.12 | Unsupervised PCA (A) and supervised PLS-DA (B) tests indicating the variable differences in clusters between the HC and Other diagnoses classes
The poor variation between the two classes may result due to the fact that the majority of the Other group of respiratory diseases included patients suffering of asthma only. These patients are normally stable, only using their inhalers occasionally. Therefore, their metabolites may result in being quite similar to those patients from the HC class.

3.3.8 | Metabolomics Analysis of Hospitals and Age Groups

PCA test between the three hospital sites BGH, GGH and PPH did not show variation in the analysis (Figure 3.13) suggesting that any different sampling approaches in each hospital did not affect our metabolomic analyses. The outlier indicated by the GGH class is sample C017 which is represented by an ex-smoker male suffering from COPD from the 70's age group that admitted himself to the hospital due to an exacerbation. The metabolites detected in his samples are displaying this fact through separation from its class.

All individuals that participated in this study have been allocated to a specific age group depending on their age. Therefore, the range of age groups started from 20s with 24 years old

as the youngest participant and finalising with the 80s with 87 years old as the oldest individual that participated. The age group classes did not show much cluster variation as it can be observed in Figure 3.13. Additionally, these results indicated that age does not necessarily overcome the disease severity, although, most patients recruited were in between 50 and 70 years old. As most studies based on respiratory diseases tend to employ elder individuals in their expertise, it is relevant to acknowledge that a more data may have given different results for the study in question.

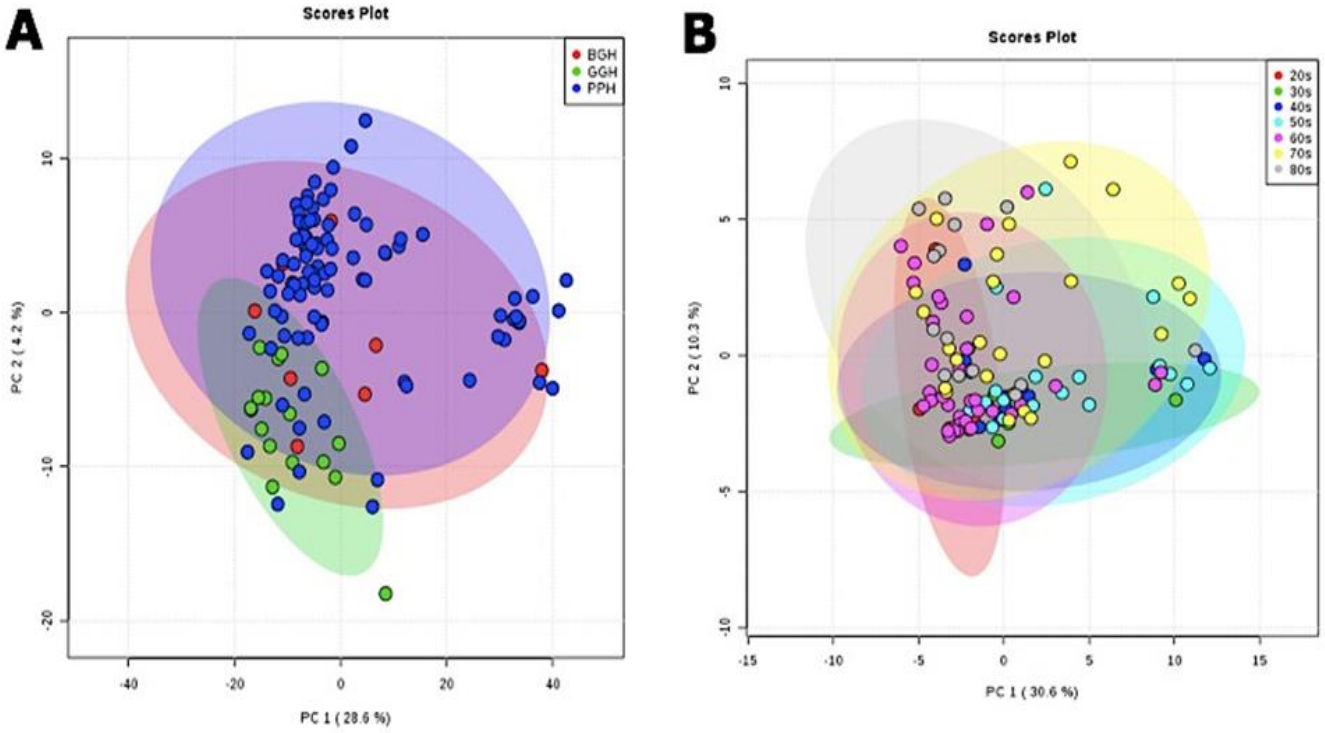


Figure 3.13 | PCA test of the three hospitals and age groups
Hospitals where the patients have been recruited and saliva samples were collected: Bronglais Hospital in Aberystwyth, Glangwili Hospital in Carmarthen and Prince Philip Hospital in Llanelli (A). The age group of every patient was recorded generally starting from patients in their 20s to individuals that were in their 80s (B).

3.3.9 | Metabolomics Analysis of COPD Severity and Gender

The PCA test based on the severity of COPD GOLD stages (previously explained in subchapter 1.2, subsection 1.2.6) indicated an accurate representation of the clusters of moderate (MO), severe (SE) and very severe (VS). The outlier of the moderate class that was located in the severe cluster was represented by sample A101 (arrowed), was a non-smoker female of the 70s age group, with an FEV1% Predicted of 66 and a medical history of asthma, acute bronchitis and pulmonary embolism.

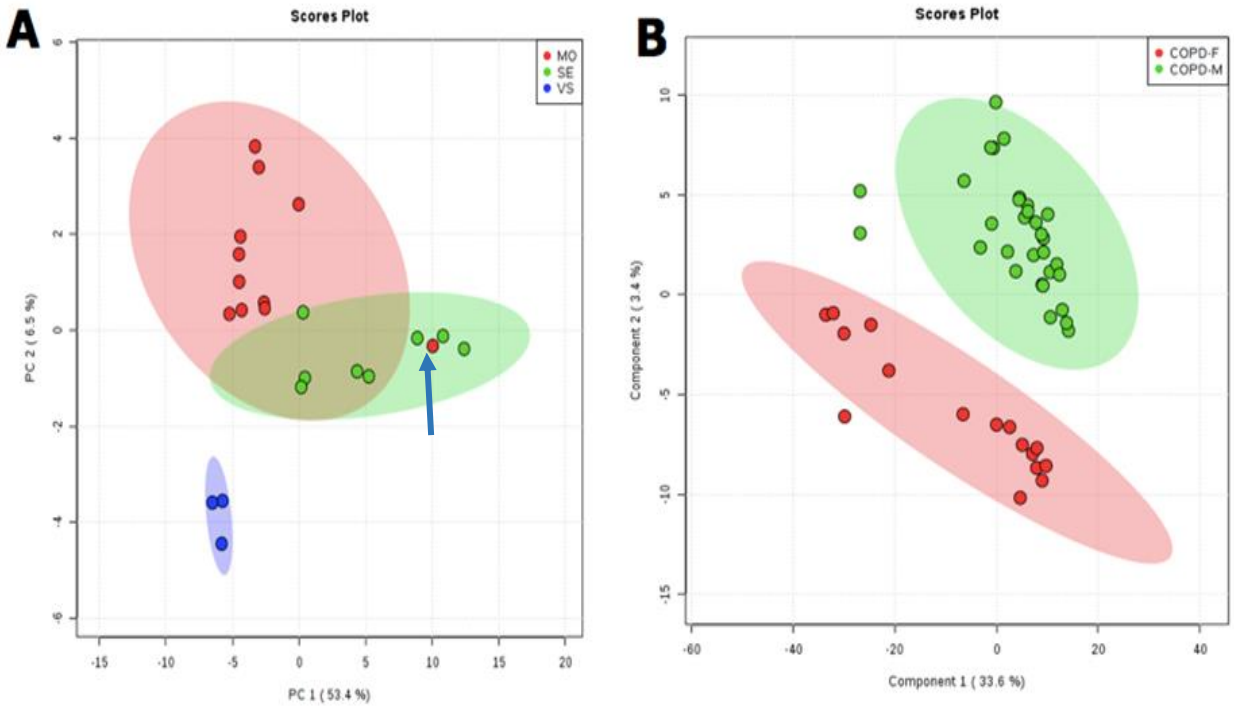


Figure 3.14 | Representation of the PCA tests of COPD severity and COPD gender
The COPD severity included moderate (MO), severe (SE) and very severe (VS) based on the FEV1% Predicted values (A). The second PCA test was performed based on the COPD gender including females (COPD-F) and males (COPD-M) as shown in B.

As for the COPD gender, there is an evident difference between the two groups of males and females. Clear separation between the clusters occurred mostly along Component 1 as indicated above in Figure 3.14.

3.3.10 | Metabolomics Analysis of Gender and Smoking Statuses

The PCA test carried out on current or ex-smokers in males and females from all four classes of COPD, LC, Other and HC (Figure 3.15A) showed a poor separation between the clusters. The outlier observed in Figure 3.15A is represented by sample C017 which was previously discussed in subsection 3.3.8.

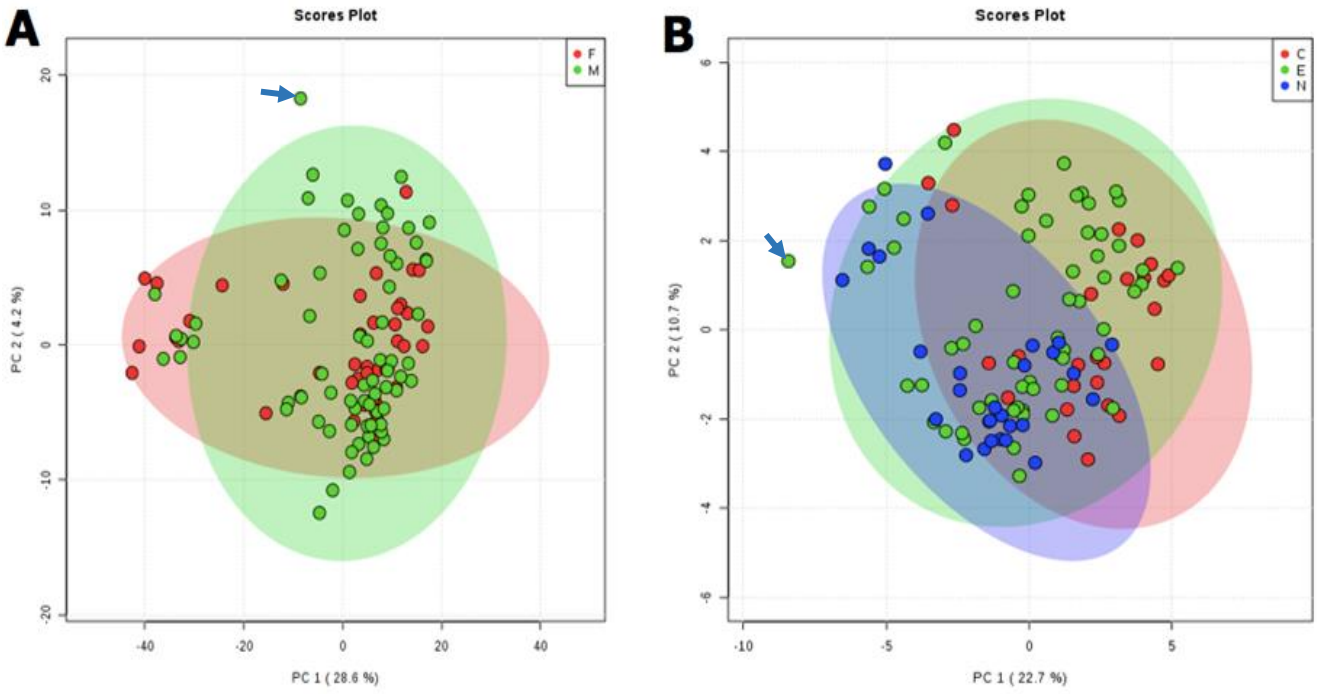


Figure 3.15 | Principal Component Analysis evaluations of the gender and smoking status of the four classes COPD, LC, Other and HC
Analysis of the current and ex-smokers did not show an effective separation between males and females (A). Although, the smoking status displayed a better clustering separation specifically with respect to the non-smokers group (B). C = current; E = ex and N = never smoked.

As for the PCA test regarding the smoking status, samples were analysed based on the Never, Ex-smoker and Current smoker grouping system. The cluster containing samples from

individuals that have never smoked showed a better separation, compared to the individuals that used to smoke or were still smoking at the time of the samples recruitment (Figure 3.15B).

The outlier detected in this case, was a sample previously discussed belonging to an ex-smoker (A067 mentioned previously in subsection 3.3.3).

3.4 | Discussion

The utilisation of flow infusion electrospray mass spectrometry for metabolomics fingerprinting employed to demonstrate if saliva allows discrimination between certain respiratory diseases and healthy controls has proved to be successful.

In this segment of work, three different classes of respiratory diseases including COPD, lung cancer, and a general group comprised of asthma, pneumonia, idiopathic pulmonary fibrosis, bronchiectasis and sarcoidosis, were compared between each other and to a group of healthy controls (never smoked, ex- and current smoker).

All PCA and PLS-DA have been generated by the analysis platform MetaboAnalyst 3.0. For all analyses data filtering, editing and normalization were performed identically, specifically to avoid introducing more variation in the four classes. In order to obtain accurate results, every analysis of the four classes was performed on the significant variables indicated by ANOVA or t-tests (depending on the number of variable grouping) with the significance plot of the P-value threshold of 0.05 (p >0.05).

Additionally, saliva metabolites from patients suffering from pulmonary diseases as COPD or lung cancer when compared to healthy controls, showed distinct separation and a good linear combination of the variables, as indicated in Figures 3.4 and 3.5.

3.4.1 | Discussion of COPD Microbiome in Comparison with LC, HC and Other

The class comprising COPD patients differentiated from LC, HC and Other classes in terms of metabolites, particularly in the supervised (PLS-DA) analyses and partial cluster separations in the unsupervised (PCA) ones. More importantly, the COPD class is the only one that contained the majority of patients post or currently on antibiotic medication for suspected or confirmed chest infections, low FEV_1% values and worrying breathlessness scores on their CAT or mMRC tests.

3.4.2 | Discussion of LC Microbiome in Comparison with COPD, HC and Other

The LC group revealed an interesting result when compared to the HC one, in the PCA particularly. The two groups overlapped and similarly to the other analyses, LC and HC always showed consistency regarding the grouping linearity. On the other hand, the PLS-DA between LC and HC, has shown a relevant separation, similarly to the analysis between LC and Other class.

It was expected for the LC class to cluster nearby the COPD or Other groups, as there was a better probability for these three classes to share similar metabolites, since their variables (Supplementary Tables 3.5, 3.6 and 3.7) coincided more. However, in relation to these results specifically, two different studies based on serum, compared COPD smokers to healthy smokers (Chen *et al*., 2015) and lung cancer to COPD patients (Deja *et al.*, 2014). In addition, *Chen et al*. (2015) discovered in their expertise a total of 23 molecular ions that were identified

in 95% sera of all study participants. These 23 metabolic markers were found to be differentially-expressed in the group of COPD smokers than in the healthy smokers. Similarly, Deja *et al.* (2014) conducted an expertise based on serum samples collected from 22 COPD and 77 lung cancer patients. By running NMR metabolic fingerprint analyses, data was then modelled using discriminant orthogonal partial least squares regression (OPLS-DA). The OPLS-DA analysis identified a number of putative biomarkers which ascertains that it is possible to successfully discriminate between COPD and lung cancer diagnoses (AUC training =0.972, AUC test =0.993). The two studies alone correlated with the results obtained in this part of work, establishing that COPD and lung cancer can be distinguished based on metabolomics biomarkers.

3.4.3 | Discussion of Hospitals and Age Group Analyses

As expected, in the analysis based on hospital sites there was no difference. All hospital classes overlapped since the diagnoses were identical. Patients almost matched between each other with regards to the study variables and disease stages. The only different feature was the geographical position. The several outliers remarked in this analysis may reflect the differences in sampling techniques and sample storage by the research nurses and research students.

In the age group analysis the separation was not evident between the age classes, although the prevailing number of patients was of those belonging to the 50s, 60s or 70s groups. However, a repeated analysis with more data for all groups may display different results. Moreover, as stated in the clinical literature and specified above in Chapter 1, the majority of pulmonary diseases develop gradually, influenced by certain factors which affect individuals at a later age.

3.4.4 | Discussion of COPD Severity and COPD Gender Analyses

The metabolomics analysis based on the COPD severity and COPD gender classes have shown the greatest degree of separations. The grouping of the COPD severity class was done based on the FEV_1% values obtained from every patients' Spirometry test results. These were moderate (MO), Severe (SE) and very severe (VS). The classification of the COPD severity was done accordingly, thus it matched both the old COPD staging as well as the new GOLD grouping (subsection 1.2.6, Tables 1.4 and 1.5). The outlier represented by a patient with MO COPD into the SE COPD class, was justified via the medical history. This patient's FEV_1% score indicated a closer value to the SE COPD group and a medical history which involved acute bronchitis, pulmonary embolism and asthma.

3.4.5 | Discussion of Gender and Smoking Status

The general analysis of the gender class was not consistent with the COPD gender analysis whatsoever. The literature falls into a disagreement regarding this subject, as there are studies which state that the males' percentage that suffer from respiratory diseases predominates over the females group. Few of the reasons are based on the fact that male smokers prevail in comparison to female smokers, and they are more exposed to health damaging factors due to the nature of their occupations (Copes, Scott, 2007). Conversely, the proportion of females suffering from pulmonary disease has increased nowadays, and has reached an even level of that of males. Some of the consequences have shown to be triggered by household air pollution and an increased number of female smokers in the last decades (Anderson *et al*., 2001; Rajer *et al*., 2014).

In the case of the smoking status analysis, the three classes of C (current smoker), E (ex-smoker), and N (never smoked) overlapped as well, suggesting there is no difference between

these three classes. Nonetheless, the N class appeared to form a closer grouping than the C or E classes.

3.4.6 | Discussion of Outliers

The outliers constituted the remarkable elements, as every class seemed to present at least one. Further investigation of each outlier has indicated that although it belonged to a certain class, its features matched more with different classes that it was attempting to join (exemplified by samples A112 and A135 HC, A067 COPD). These results indicate the impact of smoking status and medical history, the pack years and the type of life-style in general can affect a person similarly, indifferently of the diagnosis.

3.5 | Conclusions and Future Research

The Flow Infusion Electrospray Mass Spectrometry approach from this portion of work, has explored the lung metabolome with a view to investigate the differences between respiratory diseases from an easy-to-collect, minimally invasive and relatively inexpensive biofluid, saliva. Overall, all analyses have been successful or partially successful in discriminating between COPD, lung cancer and healthy patients. Moreover, with the addition of an extra group that involved other respiratory diseases such as asthma, idiopathic pulmonary fibrosis, pneumonia and bronchiectasis.

Nowadays, saliva is increasingly promoted as a suitable alternative diagnostic biofluid to blood (Patel *et al*., 2014). However, its role in pulmonary diseases is yet to be clarified. Saliva has previously been used as a biosample for metabolomics in a study involving capillary electrophoresis mass spectrometry to generate oral, breast and pancreatic cancer outlines (Sugimoto *et al*., 2010). Moreover, two studies conducted by Cameron (2015) and Cameron *et*

al. (2016), have demonstrated that saliva can deliver relevant metabolomics-derived biomarkers by employing the same metabolomics fingerprinting method of FIE-MS. This can have potential applications in the screening process for lung cancer patients.

This segment of work attempted to encourage the use of saliva for further testing, as it can have potential in determining biomarkers through metabolomics fingerprinting techniques, without detecting individual metabolites. Additionally, the FIE-MS technique offers a fast sample classification based on their Mass Spectrometry m/z values.

As an inconvenient point, this metabolomics fingerprinting method does not offer insight with regards to the biological processes that cause metabolome altering within the sample (Madsen, Lundstedt and Trygg, 2010). Therefore, the classification of disease groups is solely done on the peak masses (m/z) of samples.

As future improvement, high AUC values identified in metabolites from this study, may offer the possibility to understand aberrations that cause relevant changes in the biological pathways of metabolomes in respiratory diseases. Nevertheless, metabolomics fingerprinting techniques have evidenced discrimination between groups of respiratory diseases such as lung cancer and COPD in comparison with healthy participants.

Finally, results presented in this section of work have revealed insights into COPD, lung cancer and healthy patients, providing the ability to differentiate between their fingerprinting metabolites. Future work based on larger-scale patient cohorts, more defined diagnosis classification and a comprehensive set of all participant variables, would be required for a more in depth confirmation of the findings presented here.

CHAPTER 4 | Assessment of the Human Salivary Microbiome in Pulmonary Diseases

4.1 | Introduction

It has been highlighted that the lung microbiome does not contain a large bacterial biomass as opposed to that of the lower gastrointestinal tract, however it still shows significant variety (O'Dwyer, Dickson and Moore, 2016).

The upper respiratory tract, which comprises a considerable number of microbes, is directly linked to the lung (Wilson, Hamilos, 2014).

Dickson *et al.* (2016) state that the lung is not sterile as it was believed a few decades ago, because it is constantly exposed to microbiota through inhalation or subclinical micro aspiration from birth. Additionally, healthy lungs are affected by physiological conditions that can influence bacterial proliferation i.e. oxygen tension, local pH, temperature or inflammatory cell disruption (West, 1978; O'Dwyer, Dickson and Moore, 2016). Studies on culture-independent techniques of the healthy lung tissue (Morris *et al.*, 2013; Bassis *et al.*, 2015) mention *Bacteroidetes* and *Firmicutes* as predominant phyla.

Interestingly, chronic lung diseases (as asthma, COPD, cystic fibrosis and IPF) have a great impact on the composition of the lung microbiome as well, by altering it. Bacterial population dynamics are altered by lung diseases, as a result of changes in the immigration of microbiota from lower to upper respiratory tract and elimination achieved by cough and mucociliary clearance - processes and local conditions of the lung's microbial environment (O'Dwyer, Dickson and Moore, 2016).

The lung microbiome represents a useful source for the assessment of lung diseases' pathogenesis and inflammatory responses. The latter was tested in a study which compared inflammatory cells and microbiota identified in BAL fluid. Results indicated an increased

community abundance of *Prevotella* and *Veillonella* (belonging to phyla *Bacteroidetes* and *Firmicutes* respectively) linked to high levels of WBCs (lymphocyte and neutrophil) (Segal *et al*. 2013).

Moreover, in a study conducted on sputum samples of clinically stable COPD patients pair-sampled with samples of exacerbation periods, increases in the relative abundances of *Haemophilus*, *Pseudomonas* and *Moraxella* bacteria were found (all belonging to *Proteobacteria* phylum). These results were discovered to be in connection with exacerbation episodes (Millares *et al.*, 2014). Interestingly, Sze *et al*. (2012) in their sequencing-based tissue experiments from patients suffering of COPD, have reported an increase in the *Firmicutes* microbial community in severe COPD (GOLD stage IV) specifically. Therefore, changes in the lung microbiome due to exacerbations may be an adaptation to try to protect the lung from inflammatory responses and any potential respiratory viral infections.

The human microbiota within the organs is primarily colonized by phyla such as *Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes* and *Proteobacteria* (O'Dwyer, Dickson and Moore, 2016).

Microbial changes that are relevant for respiratory diseases are not restricted to the lung only (Aas *et al*., 2005). The human oral cavity (Figure 4.1) comprises defined spatial regions with distinct microbiomes. It is comprised of the hard and soft palates, uvula and the tonsils, gums, the tongue and saliva to coat the mouth surface. According to Dewhirst *et al.* (2010), nearly 600 microbial species have been identified within the human oral cavity. Many of which are not associated with either periodontal diseases or dental caries.

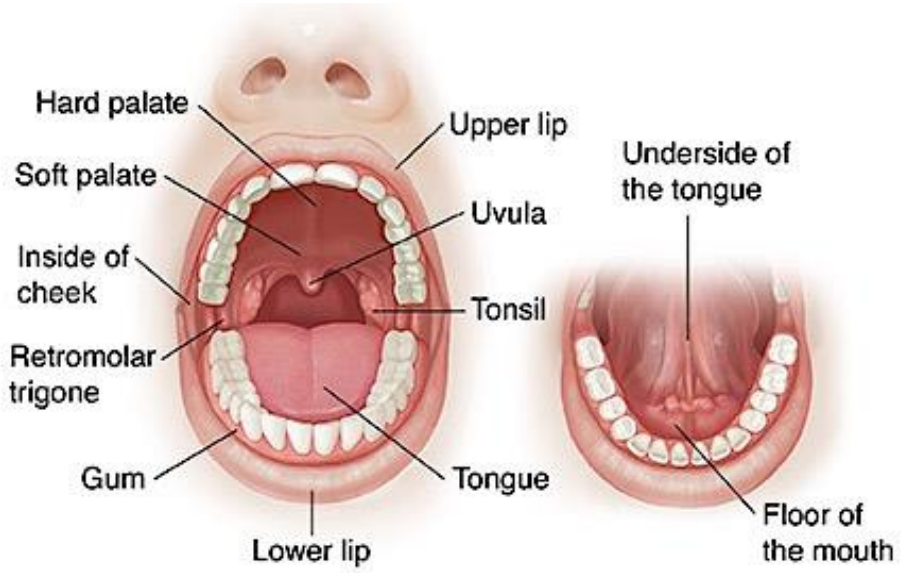


Figure 4.1 | Configuration of the Human Oral Cavity
The structure of the human oral cavity is composed of several distinct structures such as the hard and soft palates, uvula and tonsils, the gums, tongue and ultimately saliva which do not have a fixed position within the cavity. Within this structure, saliva is produced by the three salivary glands: mandibular, submandibular and parotid. Figure retrieved from Sterlingcare.com, 2017.

Periodontal diseases, for instance, which are responsible for affecting the supporting structures of teeth, can trigger exacerbations in existing comorbidities such as cardiovascular and pulmonary diseases. The constituents of the oral microbiome in periodontal diseases are capable of undergoing host dysbiosis (microbial imbalance e.g. impaired microbiota within the body), which causes the whole clinical manifestation of the disease (Liu *et al.*, 2012). Moreover, analysis of the oral microbiome of healthy patients and those suffering from periodontal disease revealed several metabolic pathways enhanced in the microbiome, which are associated with virulence factors. This resulted in a modification of the Gram-positive to a Gram-negative taxonomic domination in bacteria.

The formation of biofilms within the oral cavity, mainly on the surface of the teeth, constitutes a good example of colonising bacteria which operate as frameworks for addition of other

bacteria of the oral microbiome. According to Jenkinson and Lamont (2005), bacteria which act as frameworks include species from the *Streptococcus* and *Actinomyces* genera. These bacteria bind to salivary receptors that are able to coat uncolonised teeth surfaces. This process has been observed particularly in the *Streptococcus* bacterium which was revealed to bind to salivary proteins, such as α-amylase, fibronectin and lactoferrin (Scannapieco, 1994). Consequently, species as *Treponema* and *Candida* were found as a constant addition to the microbiome framework. Thus, the microbiome plays an important role in oral diseases for exposing causative organisms.

Saliva could be an ideal biofluid to sample the oral cavity. It does not have a set position within the oral cavity, and thus, could have microbial contributions from many parts of the upper respiratory tract. It also contains biomolecules such as enzymes, electrolytes, hormones, nucleic acids, peptides and proteins which are some of the main sources of sustenance for microbes from the oral cavity (Wickstrom and Svensater, 2008). Indeed, to exploit the diagnostic potential of the oral microbiome, Human Oral Microbe Identification Microarray (HOMIM), which consists of an oligonucleotide targeting specific 16S ribosomal Ribonucleic Acid (rRNA), has been developed. Screening of oral microbiota indicated alterations in the oral bacterial profile linked to pancreatic, oral and lung cancer, (Bonne and Wong, 2012; Huyghe *et al*., 2008, Farrell *et al*., 2012).

However, the human salivary microbiome could differentiate from the microbiomes of other oral components including the gums, tonsils or tongue. Several phyla such as *Actinobacteria*, *Bacteriodetes*, *Firmicutes*, *Fusobacteria* and *Proteobacteria* have been shown to dominate the saliva microbiome. Within these phyla, genera such as *Fusobacterium*, *Neisseria*, *Prevotella*,

Streptococcus and *Veillonella* account for the majority of the microbiome in culture-independent sequencing (Segata *et al*., 2012).

Saliva may be a promising biosample for the detection of microbiome-derived biomarkers for several diseases. Changes between the oral microbiome in a wide range of respiratory diseases, has yet to be fully determined. Work in our group has examined seasonal alterations in the saliva microbiome of healthy volunteers focusing on bacterial load and pH as well as estimated bacterial load and pH indicated significantly ($P < 0.001$) higher levels during winter and spring (Cameron *et al.,* 2015). The study also included the 16S rRNA (V3 and V4) amplicon sequencing based on DNA extracted from saliva. Results showed significant differences between participants at phylum level. The most notable pathogenic bacteria phyla detected include Actionobacteria ($P < 0.001$), Bacteriodetes ($P < 0.001$), Fusobacteria ($P < 0.001$), Proteobacteria ($P < 0.001$), Synergistetes ($P < 0.001$), Firmicutes ($P = 0.008$), and Spirochaetes ($P = 0.003$).

4.1 | Objectives

In this portion of work, DNA extracted from human saliva samples from patients that were suffering of pulmonary diseases and healthy controls, were analysed in order to observe changes in bacterial load of the human microbiome and determine their taxonomic abundances. Moreover, this study was also aimed to assess human saliva as a potential biosample that can provide information with regards to the microbiome variability in pulmonary diseases.

4.2 | Materials and Methods – Microbiomics Analysis

4.2.2 | Patient Recruitment and Saliva Sampling

The patient recruitment and saliva sampling from this section corresponds with subsection 3.2.1. Samples were centrifuged and the supernatants used for metabolomic analysis as described in Chapter 3. The pellets of the saliva samples were stored at -80 °C in the freezers of IBERS at Aberystwyth University until needed. The pellets used for DNA extractions were employed in experiments for Quantitative Polymerase Chain Reaction (qPCR) bacterial load and 16S rRNA (V3 and V4) amplicon sequencing.

4.2.3 | Processing of Raw Saliva and DNA Extraction

The protocol used was adapted from Cameron *et al*. (2015). All saliva samples donated by the participants underwent centrifugation at 10,000 x g, at 4°C for 10 minutes. Total genomic DNA was extracted from 200 µl of saliva pellet using the FastDNA Spin Kit for Soil (MP Biosciences, USA). Following the manufacturer's instructions, prior to starting the sample processing, exactly 100 ml of 100% ethanol was added to the Salt/Ethanol Wash Solution (SEWS-M) wash solution. SEWS-M is an ethanol-based wash solution used to wash away impurities once the DNA is bound to the Binding Matrix.

All saliva samples (74) were normalised to ensure 40 ng/µl DNA per sample, then centrifuged (Multifuge 3 S-R Heraeus) at 4,600 rpm at 21°C for two minutes. The normalisation process consisted in subtracting every value of quantified DNA from five, which was the "final concentration ultrapure water", in order to obtain the dilution factor. Each volume of DNA needed to add in order to make up the 40 ng/µl concentration, was obtained by multiplying the dilution factor to 40. The volume of water needed to make up the total 40 ng/µl sample was

obtained by subtracting the previously obtained volumes of needed DNA from 40. From this, 200 µl of each sample was pipetted into its corresponding Lysing Matrix E tube to which 978 µl Sodium Phosphate Buffer has been added to induce cell lysis process. Following this, 122 µl MT buffer were added to protect and solubilize the nucleic acids following cell lysis. All samples have been homogenised for 30 seconds at a speed setting of 6.0 in a FastPrep Instrument (24 MP Biomedicals, USA). Samples were then placed on ice for 30 seconds and homogenised again in the FastPrep Instrument at the same settings. These latter steps were repeated once more and after the samples were subjected to centrifugation (Thermo Scientific Heraeus Pico 21 Microcentrifuge) at 13, 000 x g for 10 minutes to pellet the debris.

The supernatant of every sample was then transferred to an individual clean 2.0 ml microcentrifuge tube to which 250 µl Protein Precipitation Solution (PPS) were added and then mixed by shaking the tube by hand 10 times. The PPS is used as a stabilization agent during sample lysis and homogenization of subsequent extraction and purification of nucleic acids. The supernatant underwent centrifugation once more at 13, 000 x g for five minutes to allow the pellet to precipitate and then transferred to a clean 5 ml Vulcan tube.

Exactly 1 ml of the re-suspended Binding Matrix suspension was added to the supernatant of each sample in the Vulcan tube. The samples were then placed on a rotator (Vortex Genie 2 Shaker) for two minutes to allow binding of DNA. The tubes were then put on a rack for three minutes to allow the silica matrix to settle. Following, 500 ml of the supernatant was removed and discarded, being careful to avoid the settled Binding Matrix.

The remaining amount of the Binding Matrix and supernatant from each tube was resuspended and approximatively 600 µl of this mixture was transferred to a Spin Filter and subjected to centrifugation at 13, 000 x g for two minutes. The catch tubes were then emptied and the

remaining mixtures from the Vulcan tubes were added to the Spin Filters and centrifuged as before. The catch tube was emptied again.

Exactly, 500 µl of the previously prepared SEWS-M wash solution, was added to the Spin Filter by gently resuspending the pellet using the force of the liquid from the pipette tip. The samples were subjected to centrifugation at 13, 000 rpm for two minutes and the catch tubes were emptied and replaced with new ones.

Without any addition of liquid, the samples were centrifuged a second time at 13, 000 x g for two minutes and then left at room temperature so that the matrix and residual wash solution can air dry. The catch tube was discarded and replaced with a new one.

The Spin Filters with the samples were left to air dry at room temperature for five more minutes. Then, the Binding Matrix from the Spin Filter was resuspended in 50 µl of DNA Elution Solution (DES) in order to wash away the unbound proteins. Finally, the samples were subjected to centrifugation at 13, 000 x g for two minutes once more, to bring eluted DNA into the clean catch tube and the Spin Filters were discarded.

4.2.4 | DNA Quantification and qPCR Standards Preparation

The resulting DNA from the extractions process was then quantified in duplicates using an Epoch Microplate Spectrophotometer (BioTek). Following the DNA quantification process, four DNA samples were chosen according to their highest values ng /µl. These samples were (concentrations and clinical diagnosis of the patient from which the sample was obtained are shown) C012/COPD – 413 ng/µl, A133/HC - 317 ng/µl, A149/HC - 230.4 ng/µl and A001/HC - 182ng/µl.

These were used to create standards that will be further used in the quantitative Polymerase Chain Reaction (qPCR) experiment.

To derive the standards, the following PCR conditions were used. Each 22.5 µl reaction volume consisted of 2.5 µl of the five neat DNA template or the non-template controls (NTC), 12.5 µl of 2 X BioMix (BioLine), along with 7 µl PCR Grade Water (Roche, Sigma Aldrich), 0.25 µl of 27f primer (5'-AGA GTT TGA TCC TGG CTC AG-3') and 0.25 µl of 1389r primer (5'-ACG GGC GGT GTG TAC AAG-3') of 10 µM concentration.

The reaction volumes were then subjected to a Thermal Cycler PCR machine (BioRad) at the following cycles: one activation cycle for 120 seconds at 94°C; 30 cycles consisting of denaturation for 45 seconds at 94°C, annealing for 45 seconds at 55°C and elongation for 90 seconds at 72°C; one cycle of final elongation step for seven minutes at 72°C and refrigeration until further processed at 4°C.

4.2.5 | Agarose Electrophoresis for DNA Extraction from Gel

After running the PCR program with the samples chosen as standards, the DNA samples were loaded onto an agarose gel electrophoresis in order to confirm the length of the 16S rRNA DNA fragment.

Exactly one gram of agarose Electrophoresis Grade (Melford Laboratories, UK) was weighted to which 100 ml of 0.5 Tris-acetate-EDTA (TAE) were added and mixed vigorously on a magnetic stirrer (Fisher Scientific, USA). Then, the mix was heated in the microwave (Panasonic NN-7543W) until the agarose melted. To this, 7.5 µl of gel stain (SYBR Safe DNA gel stain 10,000x Concentrate in DMSO Invitrogen) were added and swirled until properly mixed.

Further on, the gel tray was placed into the casting chamber and the comb was positioned into the appropriate slot in order to pour the already prepared agarose. The gel was then left to set for approximatively 30 minutes.

Exactly 2 μl of 6X Blue/Orange Loading Dye (Promega Madison, USA) were pipetted along with 5 μl of the four chosen DNA samples and a negative PCR (without template). Each dot of dye and DNA were mixed with pipette tips and then loaded into the corresponding wells of the gel. The first well was loaded with 1 μl of 1 kilobase (kb) HyperLadder (BioLine). The electrophoresis machine (BioRad Power Pac Basic) was then set for 75 minutes at 65 Volts (V).

After the electrophoresis step, the agarose gel was scanned in the Gel Doc XR machine (BioRad) under the *trans* UV light (Figure 4.2).

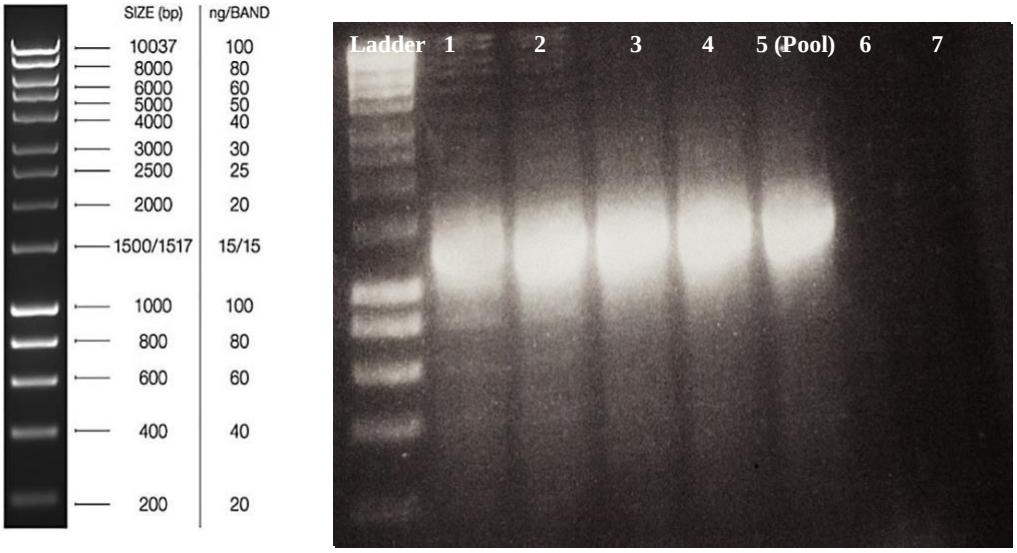


Figure 4.2 | Electrophoresis Gel Result of DNA from Human Saliva
One kb HyperLadder expressed in bp and ng/BAND on the left-hand side and then (moving to the right) each of the DNA fragments obtained from amplification of a Gene Pool mix of these four samples: C012, A133, A149 A001. The fragment is the expected ~ 1300 bp in size. A final seventh well with the PCR with negative control with no template DNA yielded no visible fragment.

Following, each band of DNA was cut from the gel under the UV light of the Transilluminator machine (DR-195M Colorado, USA). The amplified fragments were excised from a gel using a scalpel and placed in a previously weighed microcentrifuge tubes. This was then weighed once more.

The weight of the gel is directly proportional to its liquid volume and this was used to determine the amount of QX1buffer (Qiaex II Gel Extraction Kit, Qiagen, 2011) that needed to be added to dissolve the agarose (Table 4.1).

Table 4.1 | Determination of QX1 buffer amounts to extract the DNA fragment
The QX1 buffer volumes were obtained by subtracting the weight of the empty tube from the weight of the tube with the gel fragment. The resulting amounts were multiplied by 3 as stated in the Qiagen protocol (add 3 volumes for samples between 100bp – 4kb).

Lane Number	Tube with gel fragment (mg)	Empty tube (mg)	Amount of QX1 buffer to add (µl)	Number of QX1 buffer volumes	Final volume of QX1 buffer to add (µl)
1	1.24	1.11	130	3	390
2	1.30	1.12	180	3	540
3	1.34	1.13	210	3	630
4	1.37	1.12	250	3	750

Estimating that all DNA samples have in between 2 – 10 µg, exactly 30 µl of QIAEX II were added to each tube labelled 1 to 4 to bind the DNA. Afterwards, the samples were incubated in the Water Bath Machine (Clifton, UK) at 50°C for 10 minutes so the agarose can solubilize and bind the DNA. Within these 10 minutes, the samples were vortexed every two minutes in order to keep the QIAEX II in suspension. This step was repeated five times within the 10 minutes of the water bath process. The resulting colour mixture was yellow which indicated a pH of $\leq$7.5.

The samples were then centrifugated at 13, 000 rpm for 30 seconds and the supernatant was removed. The pellet was washed with 500 µl buffer QX1 and resuspended by vortexing. Then, the samples were subjected to centrifugation once more for 30 seconds and all traces of supernatant were carefully removed. This wash step was done in order to remove residual agarose contaminants. Next, the pellet was washed twice with 500 µl buffer PE and resuspended by vortexing. Then, it was centrifuged again for 30 seconds at 13,000 rpm and all traces of supernatant were carefully removed. This step was performed twice. The PE buffer prevents DNA from dissociating from the column while also serving as a useful wash to remove other components such as agarose and salt contaminants.

The pellets were then air-dried for approximately 30 minutes until the pellet became white. To elute the DNA, 20 µl of deionised water was added and the pellet was resuspended by vortexing. The resulted pellets were left at room temperature for five minutes with the lids uncapped.

Finally, the samples were centrifuged for 30 seconds at 13, 000 rpm and the supernatant was transferred into a new clean tube. This now contained the purified DNA.

Samples with extracted DNA were then quantified using an Epoch machine (BioTek, UK) and normalised to 5 ng/µl.

Following, from the gene pool values obtained by quantification the number of molecules (copies of template in the sample) were calculated from the formula:

Number of copies = $\frac{amount\ (ng)\ x\ 6.022\ x\ 10\text{^}23}{length\ (bp)\ x\ 1\ x\ 10^9 x\ 660}$ therefore, Number of copies = $\frac{59.8\ x\ 6.022\ x\ 10\text{^}23}{1342\ x\ 1\ x\ 10^9 x\ 650}$

= 4.13 x 10^{10}.

From the tube containing the extracted DNA, exactly 2 µl was transferred into a tube already containing 20 µl deionised water and mixed with the pipette tip. The sample was then vortexed and centrifuged. This sample constituted the first standard value of 10^1. From the 10^1 standard, 10-fold serial dilutions of 10^2, 10^3, 10^4, 10^5, 10^6, 10^7 were made for the qPCR experiment.

4.2.6 | 16S rRNA Quantitative PCR Set-Up

Numerous attempts at qPCR were made using different DNA and primer concentrations, agarose confirmation gels of the primer quality and different batches of SYBR Green. However, none were successful in achieving a normal amplification curve (Figure 4.3). This could be due to several factors that inhibited the efficiency of the reaction: poor primer design, suboptimal reaction conditions, pipetting errors, co-amplifications of non-specific products

(primer-dimer) and/or lack of specificity of the DNA-binding SYBR Green dye (Bio-Rad Laboratories, 2006).

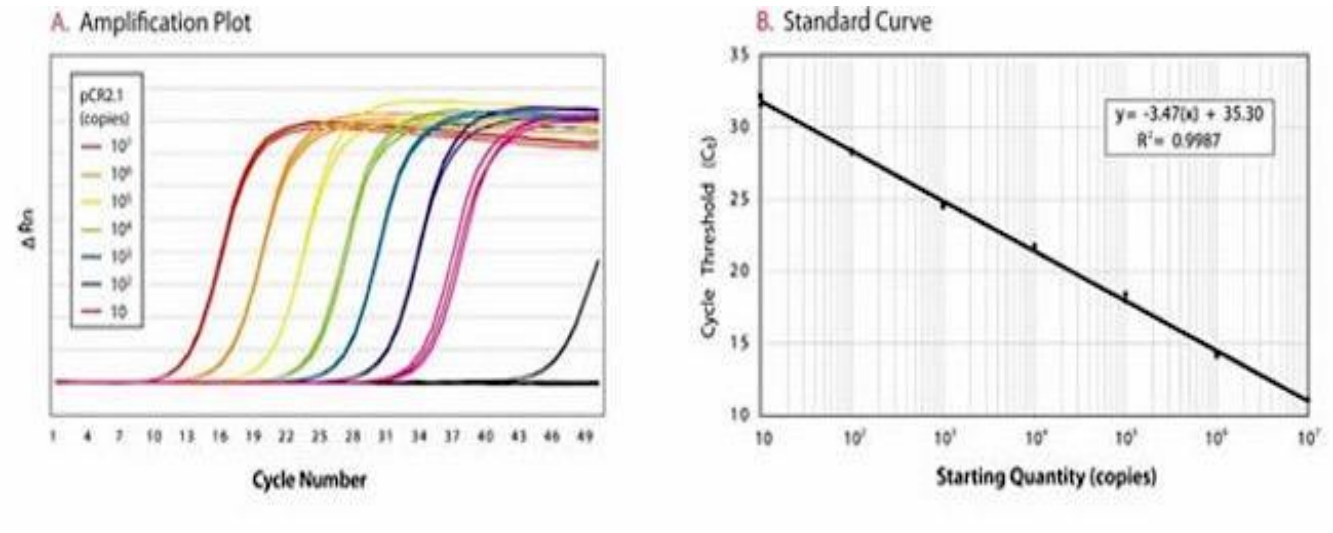


Figure 4.3 | (A) Expected amplification plot of the Quantitative PCR experiment and (B) Standard Curve generated by plotting the starting quantities (sometimes expressed as normalised Log10 values) against the Cycle Threshold (Ct). Photo retrieved from Real-Time PCR Guide, Bio Rad, 2006.

Finally, the standard DNA samples were extracted from a whole genomic *Escherichia coli* (*E. coli*) culture which gave the expected results.

The qPCR experiment was prepared under a fume cupboard in a 96-well plate in duplicates for each sample including standards and no template controls (NTC). Each reaction volume of 11.28 µl consisted of 5 µl SYBR Green the DNA binding dye, 0.04 µl EubF1 primer (5'-GTG STG CAY GGY TGT CGT CA-3') and 0.04 µl EubR1 (5'-ACG TCR TCC MC ACCT TCC TC-3') both of 10 µM concentration, 3.20 µl PCR Grade Water and instead of adding 1 µl of DNA as in the previous experiments, 3 µl were added. For the NTC sample, the DNA was replaced with PCR Grade Water.

The qPCR machine used was Light Cycler 480 II Roche, under the following settings: one cycle of activation for 10 minutes at 95°C, 40 cycles of denaturation for 15 seconds at 95°C,

40 cycles of annealing for 60 seconds at 60°C, followed by a melt curve consisting of a temperature gradient of 60 to 95°C in 0.5°C increments, each for five seconds.

4.2.7 | 16S rRNA Quantitative PCR Data Analysis

Estimates of the bacterial load from saliva were calculated through qPCR, using the 10-fold serial dilutions of 16S rRNA gene standards.

The qPCR cycle number is displayed on the x-axis and the fluorescence detected during the amplification reaction, on the y-axis. The fluorescence is directly proportional to the amount of amplified product; therefore, it gets measured in each cycle. The threshold cycle (C_T) is the cycle number at which detectable signal is achieved. It is determined by the amount of template present at the start of the amplification reaction (Bio-Rad Laboratories, 2006).

In this experiment, results indicated a variety of both high and low C_Ts as shown in Figure 4.5.

The standard curve for the bacterial load was constructed by plotting the log of the dilution factor of the starting quantities against the C_T values from the amplification of each dilution.

From this, the equation of linear regression line, the coefficient correlation (R^2) and the slope (m) were calculated from the following: $y = mx + b$,

resulting in $y = -3.5862x + 42.933$ with a coefficient correlation of $R^2 = 0.9154$.

By knowing the slope value, the efficiency (E %) was determined from the following:

$\%E = 10^{-(1/slope)}$,

therefore $10^{-(1/-3.5862)} = 1.9$ and

$\%E = (1.9 - 1) \times 100 = 90\%$ efficiency

4.2.8 | 16S rRNA Amplicon Sequencing Set-Up

Sequencing of the 16S rRNA gene was carried out through amplification of the V3 and V4 region. Following, amplicon sequencing on the Illumina MiSeq instrument was performed. Initially, the V3 and V4 region of the 16S rRNA gene was amplified via duplicate PCR with locus specific primers. PCR Grade Water was utilised as negative control.

Each reaction of 25 µl volume, consisted of 12.5 µl 2X Accuzyme Mix (BioLine), 5 µl of 806r PCR Primer (5'-GAC TAC HVG GGT ATC TAA TCC-3' with Illumina reverse overhang adapter sequence 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-3'), 319f PCR Primer (5'-CCT ACG GGN GGC WGC AG-3' with Illumina forward overhang adapter sequence 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG-3') each of 1 µM concentration and 2.5 µl normalised DNA as detailed by Klindworth *et al.*, (2013).

The V3 and V4 are two of the nine "hypervariable regions" of the bacterial 16S rRNA, which have been shown to possess significant sequence diversity among certain bacteria (Chakravorty *et al.*, 2007). These two hypervariable regions of 16S rRNA were employed in phylogenetic classifications of the microbial populations within human saliva. The overhang adapter sequences were added to the 806r and 319f primer pair sequences for compatibility with the Illumina index (designed by Cameron, 2015) displayed in Supplementary Table 4.1.

Reactions were run under the following PCR conditions: one cycle activation for 10 minutes at 95°C, followed by 25 cycles of denaturation for 30 seconds at 95°C, annealing for 30 seconds at 55°C, elongation for another 30 seconds at 72°C, and five minutes for final elongation at 72°C as well, then held at 4°C. Each PCR product was confirmed through visualisation on a 1% agarose gel, that was run for 90 minutes at 70 V.

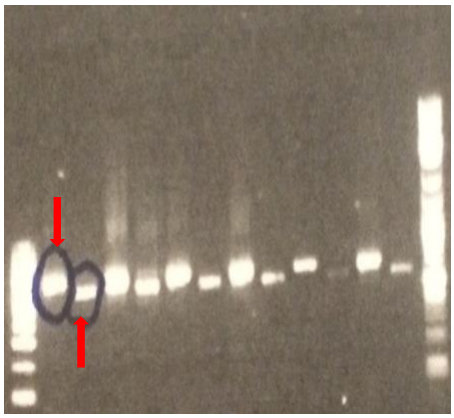


Figure 4.4 | AMPure clean-up to check the change in size of the DNA sequence after adding the Illumina adaptors, 1st round sample as given initially and 2nd round Illumina adaptors.

Following confirmation of accurate PCR results, the reaction volumes were purified using the AMPure beads (Agencourt Qiagen) clean-up to remove adapter dimers and confirmed on an agarose gel once more (Figure 4.4).

These further steps were performed with the help of Dr Matthew Hegarty - IBERS Translational Genomics Facility Gogerddan. After the dimer purification step, one last PCR program was run to attach Illumina adaptors to amplified products. This step allows multiplexed amplicon sequencing on the Illumina MiSeq platform. The final reaction volume of 25 µl, constituted of 2.5 µl of the purified PCR product obtained in the previous step, 12.5 µl of 2X Accuzyme Mix (BioLine), 5 µl of PCR Grade Water (Roche), 2.5 µl of the relevant Nextera XT Index Primer1 (N7##) and Nextera XT Index Primer2 (S5##) from Illumina (Supplementary Table 4.1) (Cameron, 2015). Subsequently, the reaction mix was subjected to a PCR program consisting of three minutes at 95°C, eight cycles of 30 seconds at 95°C and 30 seconds at 55°C, followed by a final elongation step of five minutes at 72°C. Following, the entire reaction volume was verified on a 1% agarose gel, the PCR products excised from it and

purified using an Isolate II PCR and Gel Extraction kit (BioLine) following manufacturer's instructions.

4.2.9 | 16S rRNA Amplicon Sequencing Data Analysis

Further processing of samples was carried out at the IBERS Translational Genomics Facility Gogerddan from Aberystwyth University with the help of Dr Matthew Hegarty. The sample libraries obtained were pooled together in equimolar concentration. Following, these were sequenced with 20% PhiX DNA as control for low diversity on the Illumina MiSeq platform using MiSeq V3 reagents for a 2 x 300 bp paired end sequencing.

After sequencing, the reads were merged and compressed, then subjected to the MG-RAST and European Nucleotide Archive (ENA) pipelines for metagenomic analysis (Meyer *et al*., 2008).

4.3 | Results

4.3.1 | 16S rRNA Quantitative PCR

Initial studies sought to test if there were any changes in bacterial load of the saliva samples from respiratory disease patients by qPCR. The standards curve obtained derived producing an almost perfect doubling for each amplification cycle and a linear standard curve, confirming the 10-fold serial dilution of DNA with an R^2 confirms a value of 0.915 (Figure 4.5). Subsequent results obtained from the qPCR analysis of the saliva samples were transferred to a Microsoft Excel (2013) spreadsheet. The 16S rRNA bacterial loads within the unknown DNA

samples were calculated by constructing an absolute standard curve against the known standards of the *E. coli* DNA concentrations.

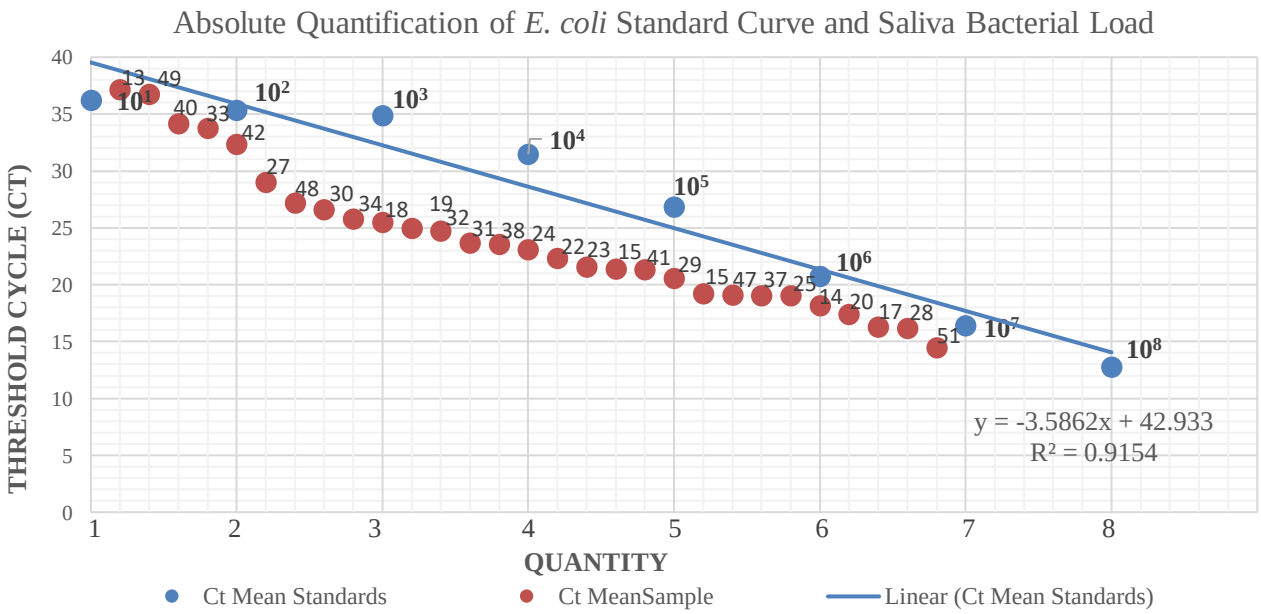


Figure 4.5 | Absolute quantification of *E.coli* standard curve for reaction optimization of unknown bacterial load in DNA from saliva samples
Standard curve generated for the quantities of 16S rRNA gene copies (bacterial load) from saliva samples were obtained by plotting the log of the starting quantity template of *E. coli* DNA against the C_T values from the amplification of each unknown saliva sample. The equation of the regression line and the coefficient correlation (R^2) are displayed on the graph above. The calculated amplification efficiency was 90%.

Table 4.2 | Quantitative PCR sample order of the Absolute Quantification Analysis containing Sample ID, Diagnosis Class, Smoking Status and Pack Years recorded per patient during recruitment

Samples in decreasing order as displayed in Figure 4.5, starting from the highest concentration of the absolute quantification, to the lowest concentration of 10^8.

qPCR Sample Order	Sample ID	Group (Diagnosis)	Smoking Status	Pack Years	Suspected Infection Yes/No
S13	A055	COPD	current	35	No
S49	A125	HC	ex-smoker	20	No
S40	A115	COPD	ex-smoker	53	No
S33	A095	HC	never	N/A	No
S42	A117	HC	ex-smoker	1.75	No
S27	A088	HC	never	N/A	No
S48	A124	HC	ex-smoker	6.25	No
S30	A091	HC	never	N/A	Yes
S34	A101	COPD	never	N/A	Yes
S18	A063	COPD	ex-smoker	44	No
S19	A067	COPD	ex-smoker	28	Yes
S32	A094	LC	current	40	Yes
S31	A093	HC	never	-	No
S38	A112	COPD	current	45	Yes
S24	A079	HC	never	-	No
S22	A077	COPD	ex-smoker	38	Yes
S23	A078	HC	never	-	No
S15	A059	HC	current	80	Yes
S41	A116	COPD	ex-smoker	80	No
S29	A090	LC	ex-smoker	30	No
S47	A061	COPD	ex-smoker	60	Yes
S37	A123	HC	current	13	Yes
S25	A110	COPD	ex-smoker	50	Yes
S14	A086	HC	never	-	No
S20	A058	COPD	current	30	No
S17	A071	COPD	current	50	Yes
S28	A062	COPD	ex-smoker	45	Yes
S51	A089	COPD	ex-smoker	20	Yes

Based on the derived standard curve the bacterial load for each experimental class was calculated (Figure 4.6).

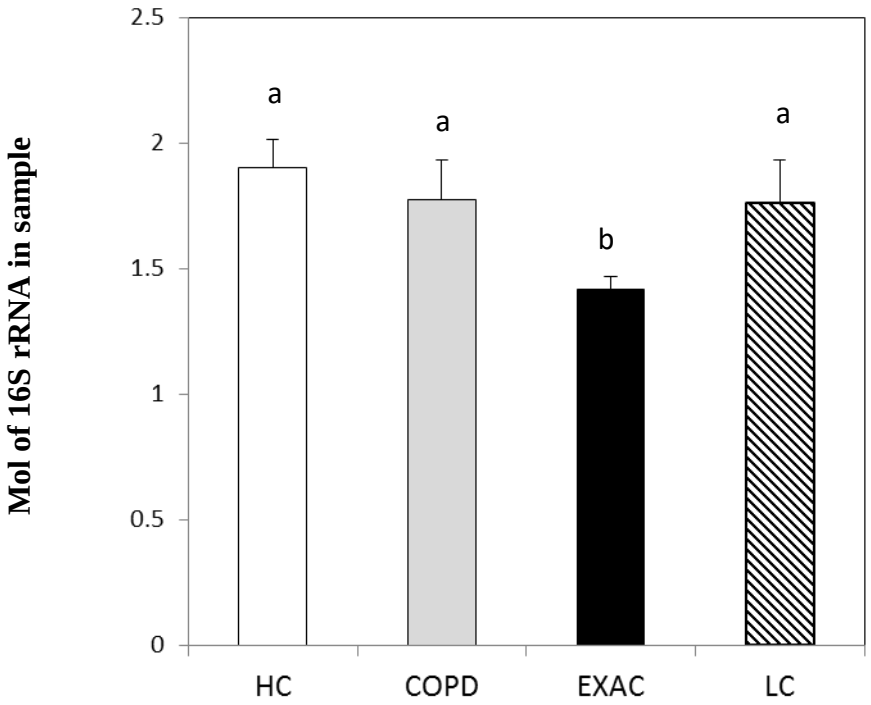


Figure 4.6 | Bacterial load in DNA from saliva samples from respiratory patients and healthy controls
Mean bacterial loads for healthy controls (HC) and patients with base-line chronic obstructive pulomonary disease (COPD), COPD patients which were undergoing an exacerbation (EXAC) and lung cancer (LC) patients. Significance groups ($P < 0.05$) as established by ANOVA are indicated by letters.

Analyses of bacterial loads in saliva indicated that the samples obtained from COPD patients which were undergoing an exacerbation (EXAC) had significantly lower loads than any other category. Saliva from healthy controls (HC), non-exacerbating COPD patients (COPD) and lung cancer patients did not significantly differ.

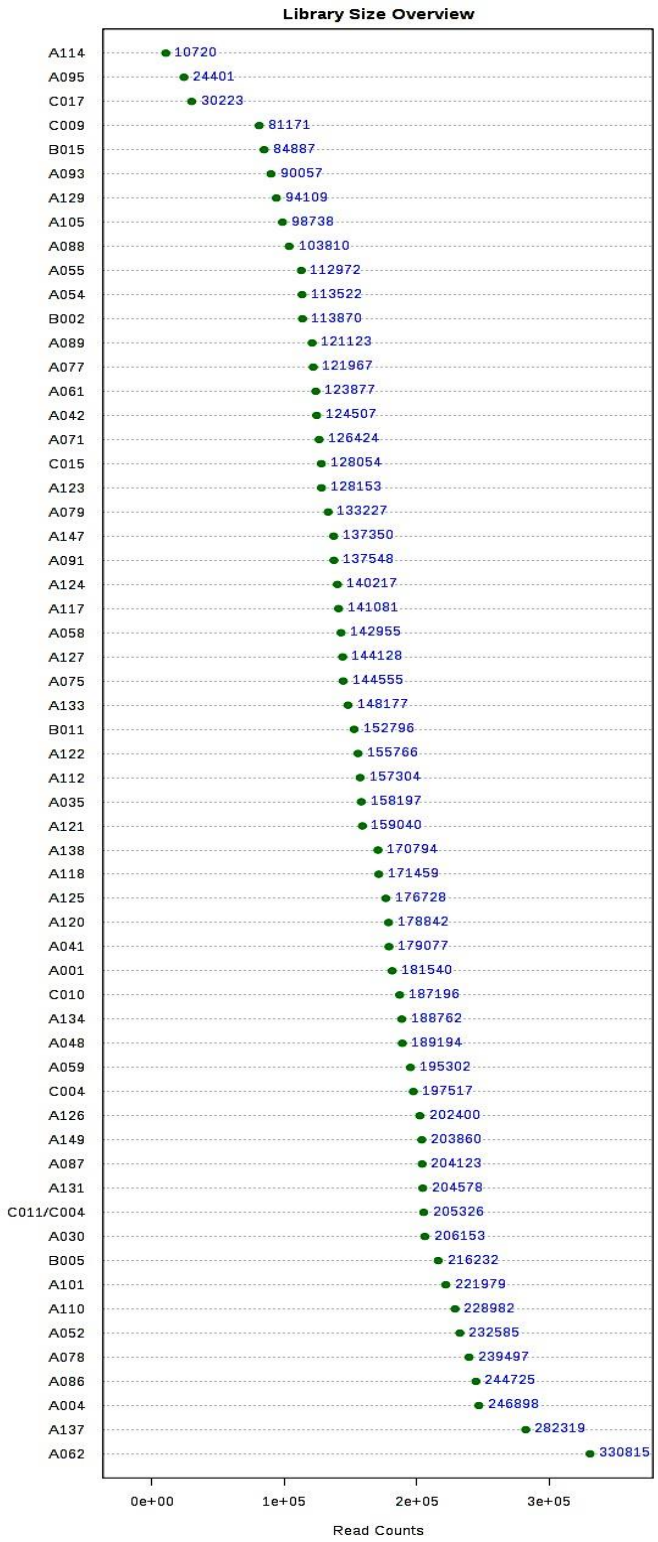
4.3.2 | 16S rRNA Amplicon Sequencing Results

The following analyses were carried out with the help of my supervisor Luis Mur – IBERS, Aberystwyth University.

The V3 to V4 regions of 16S rRNA genes of the microbial genomes of the DNA samples were sequenced using the MiSeq Illumina platform following a paired end approach. Derived sequences were assembled and analysed against the SILVA (High Quality ribosomal RNA) database (Quast *et al*., 2013) to identify the derived Operational Taxonomic Units (OTUs) using the European Nucleotide Archive (ENA) pipelines (Meyer *et al.*, 2008).

The number of paired reads for each sample is given in Figure 4.7. Subsequently, the matrix generated by the ENA platform was interrogated using the Microbiome Analyst Platform (Dhariwal *et al*., 2017).

Initial analyses focused on establishing if there was a core microbiome in all saliva samples (Figure 4.8). Only Streptococcus was prevalent in every saliva sample. The data were next assessed for any differences that could be linked to the different experimental classes (Figure 4.9).



Sample ID	Sample Diagnosis
A114	HC
A095	HC
C017	COPD (exac.)
C009	COPD (exac.)
B015	COPD
A093	HC
A129	HC
A105	COPD
A088	HC
A055	COPD
A054	COPD
B002	COPD
A089	COPD
A077	COPD
A061	COPD
A042	COPD
A071	COPD
C015	COPD (exac.)
A123	HC
A079	HC
A147	HC
A091	HC
A124	HC
A117	HC
A058	COPD
A127	HC
A075	HC
A133	HC
B011	COPD
A122	HC
A112	HC
A035	COPD
A121	HC
A138	HC
A118	HC
A125	HC
A120	HC
A041	COPD
A001	HC
C010	COPD (exac.)
A134	HC
A048	COPD
A059	HC
C004	COPD (exac.)
A126	HC
A149	HC
A087	HC
A131	HC
C011/C004	COPD (Post-exac.)
A030	HC
B005	COPD
A101	COPD
A110	COPD
A052	HC
A078	HC
A086	HC
A004	HC
A137	HC
A062	COPD

Figure 4.7 | Pair-end reads for each saliva sample with their corresponding Sample ID and Diagnoses

This step classified the groups of the Operational Taxonomic Units (OTUs) according to the 16S Bacterial genotypes identified in the saliva samples.

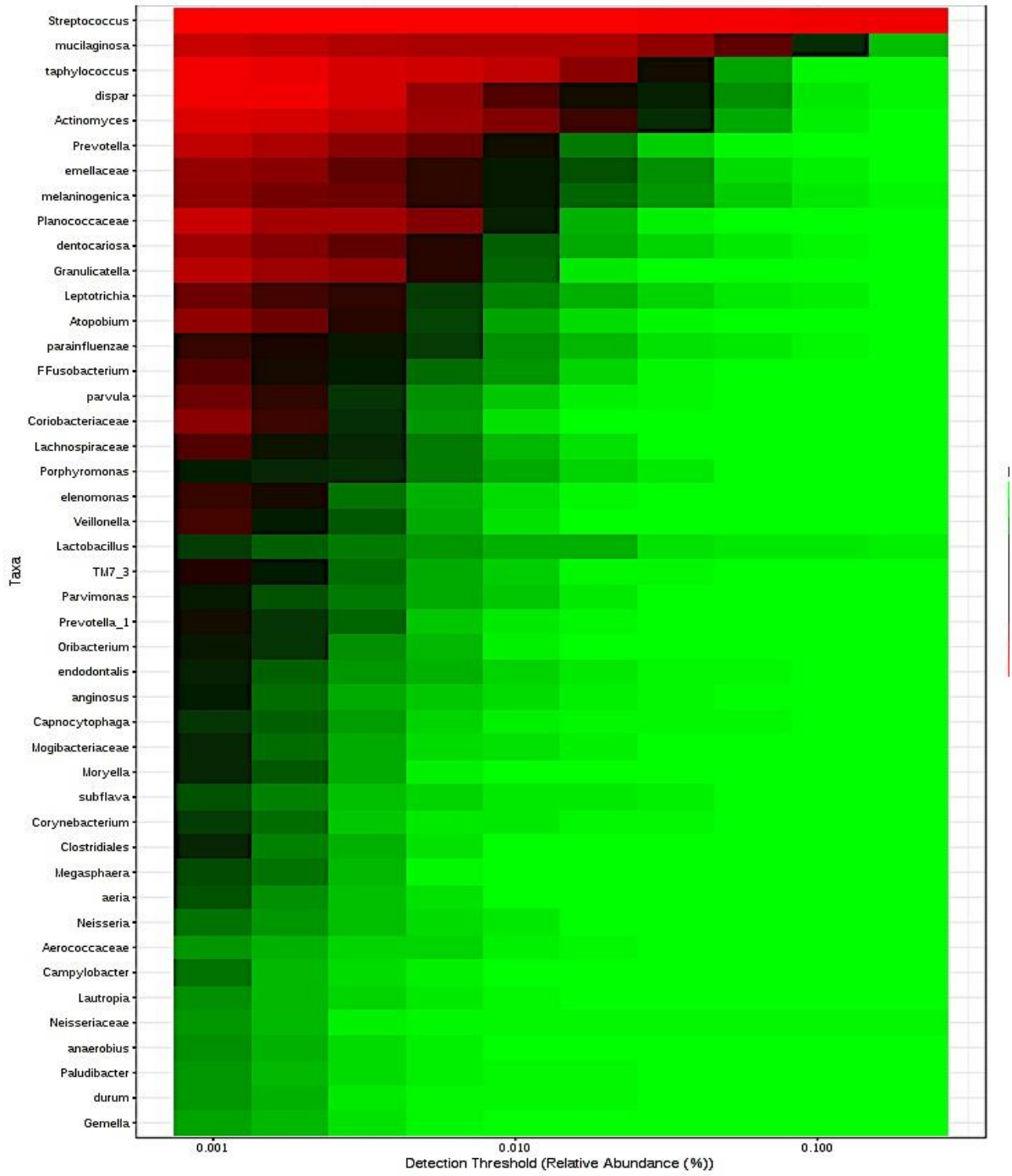


Figure 4.8 | Prevalence of OTUs in the saliva dataset
Each block of colour represents 10% of the microbiome data analysed.

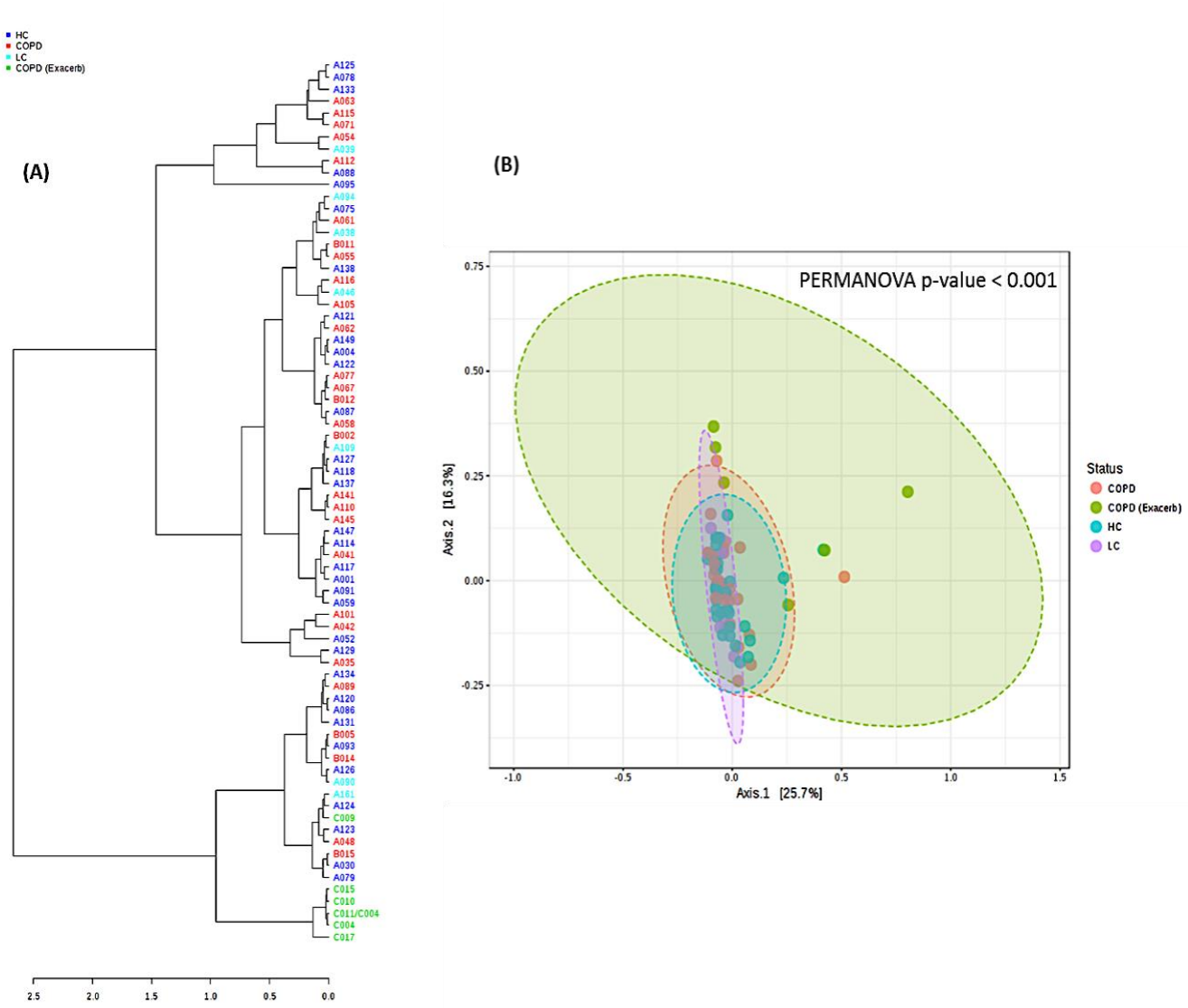
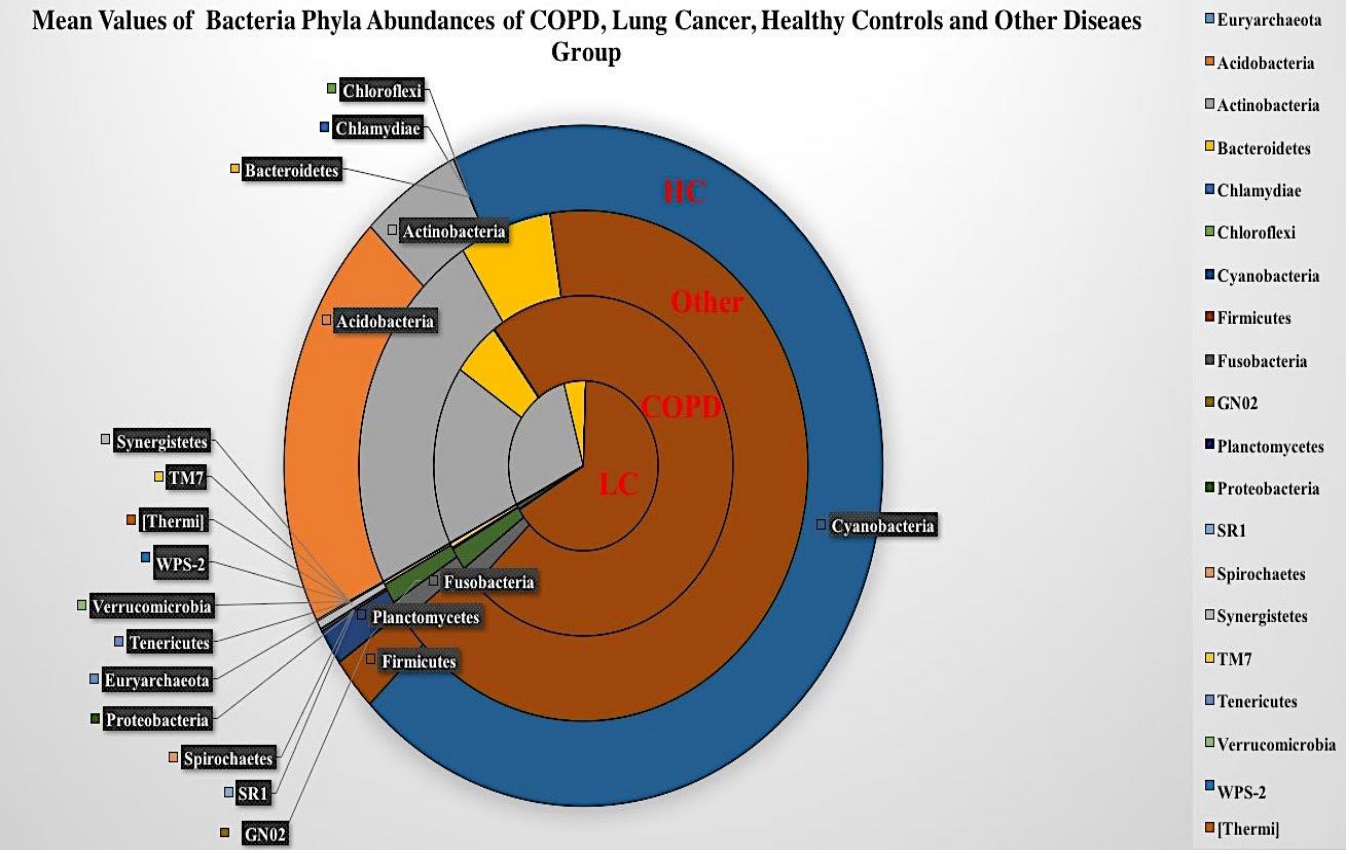


Figure 4.9 | Assessing variation in the microbiome linked to experimental classes and displayed as a (A) dendrogram and (B) Principal Coordinate Analysis (PCoA) with estimates of significance derived by PERMANOVA

Permutational multivariate analysis of variance compared COPD, COPD (Exac), HC and LC groups and if the centroids and dispersion of the groups are significantly different.

Figure 4.10 | Classification of the means of phyla taxonomy for 16S rRNA sequencing of the four classes COPD, LC, HC and Other

Individual variation of the bacteria phyla abundances was identified in the DNA samples from human saliva, through the European Nucleotide Archive (Meyer *et al*., 2008) platform. Bacterial phyla displayed significant different abundances between the four groups of COPD, Healthy Control, Lung Cancer and the Other group.



This was also clearly indicated by a significant reduction in the *a*-biodiversity in the COPD exacerbator group. Alpha diversity captures both the organismal richness of a sample and the evenness of the abundance distribution. Similar observations were made when analysing at the genus (Figure 4.11) and species (Figure 4.12) classification levels.

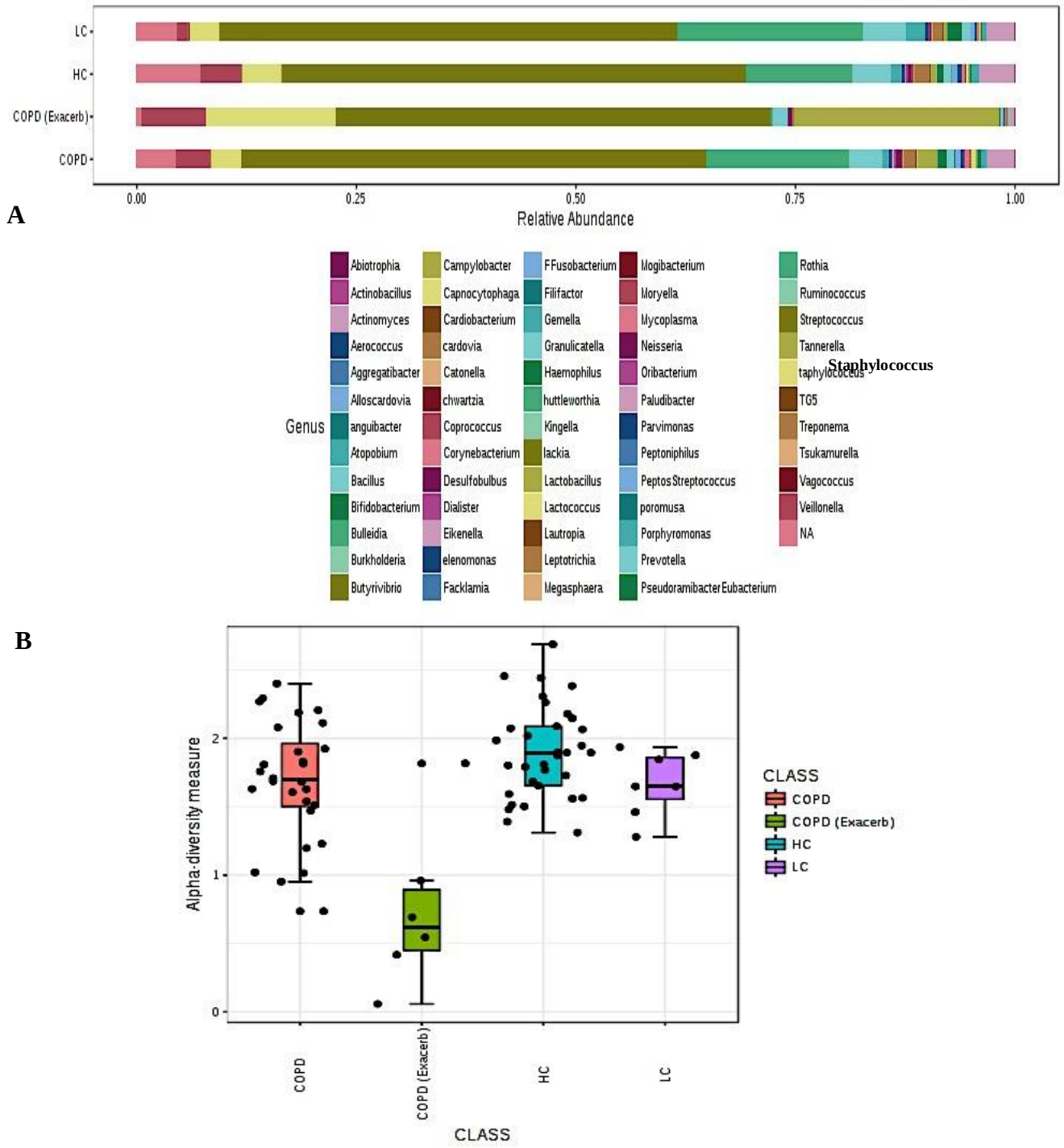


Figure 4.11 | Sample classification based on the means of genus taxonomies for 16S rRNA sequencing of the four classes COPD, COPD Exacerbation, LC and HC

(A) Variation of the bacterial genomic class abundances was identified in the DNA samples from human saliva, through the European Nucleotide Archive platform (Meyer *et al.*, 2008). Bacterial genera displayed significant different abundances between the four groups of COPD, COPD Exacerbation, Healthy Control and Lung Cancer. **(B)** The α-diversity of each experimental class.

The lowest classification level OTUs could be identified (as shown by the low number of NA's = no OUT abundance threshold applied) were at the genus level; differences between the experimental classes were plotted (Figure 4.13).

PCA of the genus level further indicated that the COPD exacerbator group was distinctive (Figure 4.13A). ANOVA of the genus level identified 11 genera which significantly differed between the experimental classes. These were plotted in a heat map for individual samples (Figure 4.13B) and as the average for each experimental class (Figure 4.13C). These analyses clearly indicated that the COPD exacerbation stage was associated with a loss in a range of bacterial genera but an increase in *Staphylococcus*.

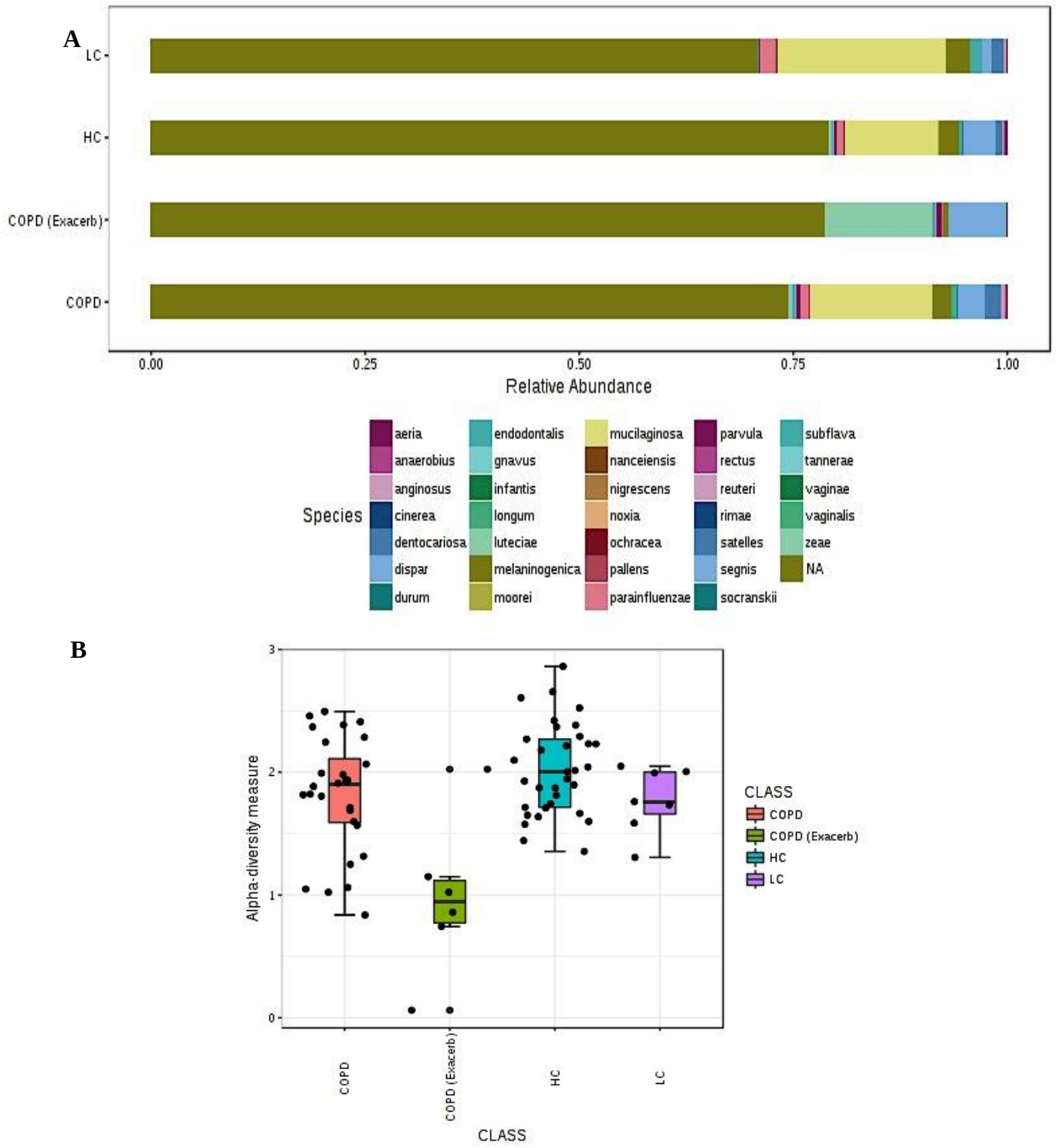


Figure 4.12 | Classification of the means of species taxonomy for 16S rRNA sequencing of the four classes COPD, COPD Exacerbation and HC

(A) Individual variation of the bacteria species abundances was identified in the DNA samples from human saliva, through the European Nucleotide Archive (Meyer *et al.*, 2008) platform. Bacterial species displayed significant different abundances between the four groups of COPD, COPD Exacerbation, Healthy Control and Lung Cancer. **(B)** The a-diversity of each experimental class.

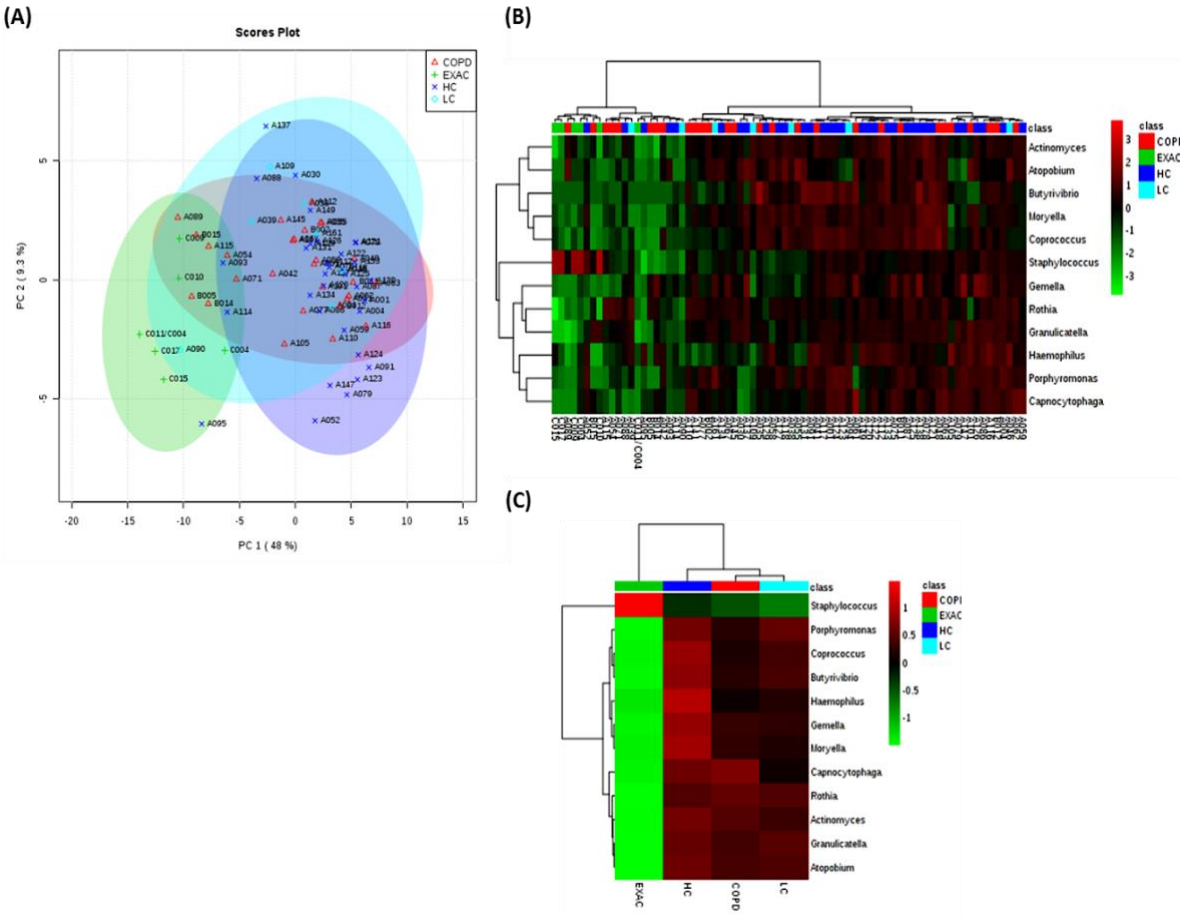


Figure 4.13 | Identifying genera which differed according to experimental class COPD, LC, HC and COPD Exacerbation (EXAC) in human saliva
(A) PCA of genera abundance **(B)** Plot of a heat map and dendrogram of genera showing significant ($P < 0.05$) variation between the classes as defined by ANOVA and plotted for each sample **(C)** Heat map and dendrogram of the averages of genera showing significant ($P < 0.05$) variation between the classes as defined by ANOVA.

To demonstrate the variation in individual genera amongst the different experimental classes these were plotted using a box and whisker analysis (Figure 4.14).

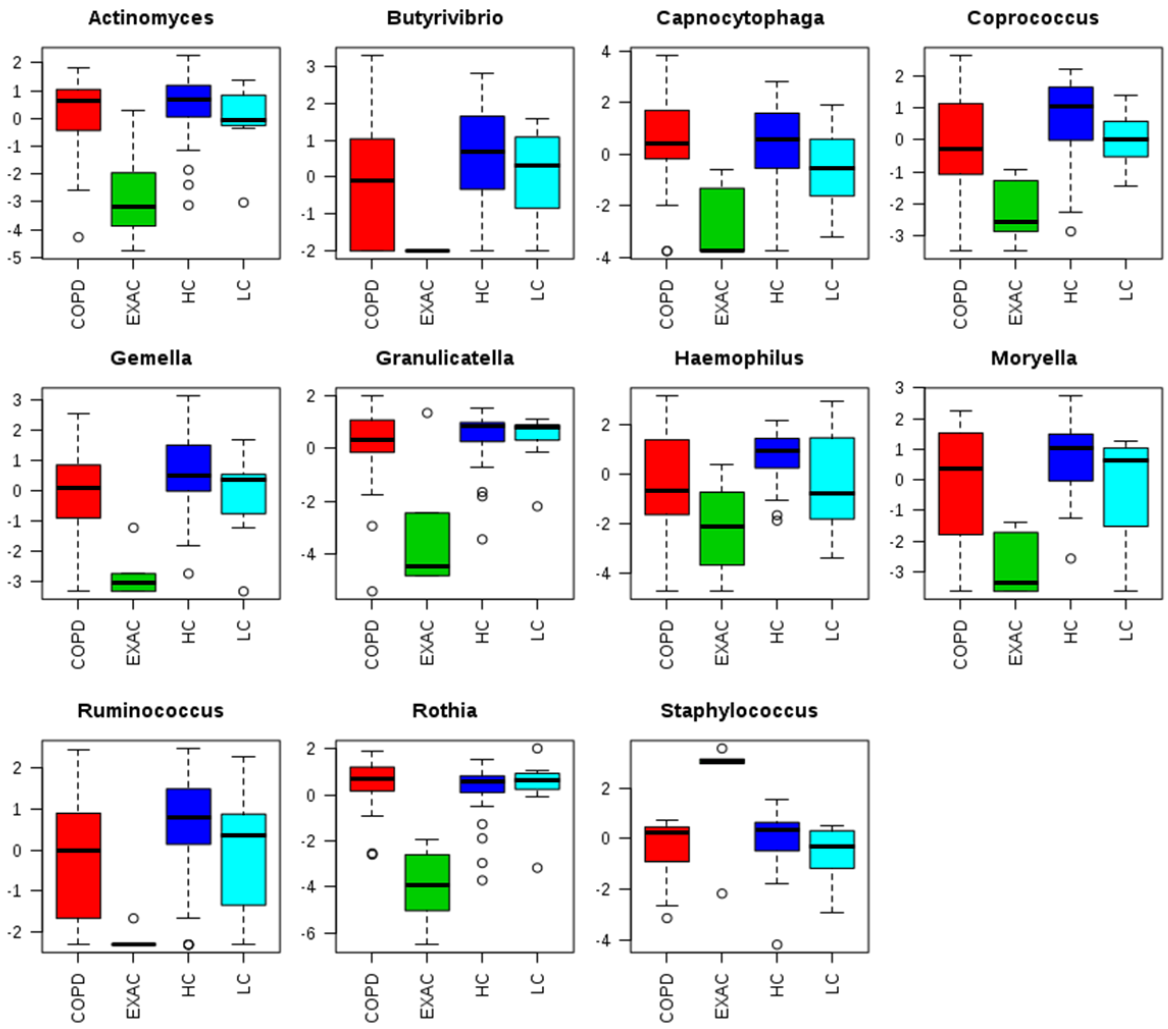


Figure 4.14 | Plots of genera that showed significant (P <0.05) variation between the experimental classes as defined by ANOVA

Figure 4.14 confirmed that in the vast majority of genera, they were significantly reduced at the exacerbation stage (in the COPD group), the one exception being *Staphylococcus*. Plotting the species of *Staphylococcus* that were detected in the saliva microbiome, this increase appeared to be linked to changes in *S. aureus* which was by far the most abundant *Staphylococcus* species in the saliva samples (Table 4.3).

Table 4.3 | Percentage of *Staphylococcus* species in each experimental class

Group	*Aureus*	*Epidermidis*	*Haemolyticus*	*Equorum*	*Pettenkoferi*	*Sciuri*	*Succinus*
COPD	99.76666	0.211947	0.020741	0.01	0.01	0.000648	0.01
EXAC	99.98756	0.001422	0.000356	0	0.000711	0.005511	0.004445
HC	99.90245	0.088344	0.006699	0.000419	0	0.002093	0.01
LC	99.98037	0.016361	0.003272	0.01	0.01	0.01	0.01

S. aureus is frequently found in the nose, respiratory tract, and on the skin. *S. aureus* is an important cause of pneumonia, but in upper respiratory tract often exists as a 'pathobiont' in that they colonize the host asymptomatically. Such pathobionts include *Streptococcus pneumoniae*, *Haemophilus influenza*, as well as *Staphylococcus aureus*. However, these can still cause disease as people age, are hospitalised or are smokers (Schenck *et al.*, 2011). It is of relevance that COPD is a risk factor in initiating *S. aureus* –mediated pneumonia (Wooten *et al.*, 2013). However, increases in *S. aureus* – changes have not been previously described in the saliva of COPD patients. Increases in *S. aureus* may reflect interactions with other species such as the fungus *Candida albicans*. *S. aureus* is known to form multispecies biofilms with *C. albicans* to help penetrate human tissue (Schlecht *et al.*, 2015).

The loss in bacterial diversity is a major feature of COPD exacerbation posing the question of why this should be the case.

The following is a summary of some of the features of the bacterial genera which differentially accumulated between the COPD exacerbator and other classes.

Some genera include opportunistic pathogens as Actinomyces, which are commonly found in the gums and can lead to the opportunistic infections of the oral cavity and rarely cause Actinomycosis which leads to the formation of abscesses in the mouth or lungs; *Capnocytophaga* is a commensal species common to the oropharyngeal tract of mammals. However, it can be involved in infections depending on the immune status of the patient. In

immunocompetent patients, it could contribute to periodontal infections occasionally leading to apical and periodontal abscesses (McGuire & Nunn, 1996).

Gemella are found in the mucous membranes of airways and have been linked to exacerbations in cystic fibrosis patients (Carmody *et al.,* 2013).

Granulicatella genus is known to be a normal flora of the upper respiratory and gastrointestinal. It has been linked to diseases like Endocarditis and Bacteremia, but it is noted that in lungs it increases in the sputum of LC patients (Cameron *et al.,* 2017). A similar increase in the saliva of LC patients was not observed.

Other bacterial genera appear more benign. *Rothia* is considered a part of the normal flora of the human oropharynx and upper respiratory tract and not associated with disease. *Botyvibrio* is commonly associated with the gastro-intestinal tract (GI) tract but have been noted to occur in the mouth (Segata *et al.,* 2012). However, it has not been associated with a respiratory tract pathology.

Coprococcus has also been linked to maintaining gastro-intestinal health but there is no evidence any role in the oral cavity (Backhed *et al.,* 2015). Similarly, *Moryella* and *Ruminococcus* has not been previously associated with the oral cavity.

Considering the data together, it can be seen that reduction in the numbers of certain genera cannot be linked to any abilities to act as an opportunistic pathogen or respond to host immune status. Instead, it seems likely that this could reflect patients being treated with antibiotics. This would also agree with the reduction in overall bacterial load seen in the COPD exacerbator group. Given that *S. aureus* is breaking this trend it may be that antibiotic treatments are selecting for resistant bacterial strains of a significant pathogen in the oral cavity.

S. aureus is not the only species to increase in the exacerbator category. Although species data are limited (Figure 4.12), there is a significant increase in *Lactobacillus zeae*; a subspecies of

Lactobacillus casei. This species is a facultatively anaerobic species and possess a strictly fermentative metabolism with lactic acid as the major metabolic end product (Axelsson, 1998). This would suggest a change to anaerobic respiration in the upper respiratory tract of COPD patients suffering of exacerbations towards the increased prevalence of bacterial species which are tolerant to higher levels of lactic acid.

4.3.2 | COPD Stage Specific Microbiome Changes

Given the observed difference between the non-exacerbator and exacerbator categories (e.g. Figure 4.14), the salivary microbiomes were assessed for differences that aligned with COPD stages; MILD (MI), MODERATE (MOD), SEVERE (SEV) and VERY SEVERE (VS), EXACERBATOR (EXAC) and HEALTHY CONTROL (HC) (Figure 4.15).

PCA of genera abundance suggested that the EXAC category remained the most distinctive (Figure 4.15A). ANOVA identified genera which significantly differed between the COPD categories and when displayed on a heat map (Figure 4.15B), this also indicated that only the EXAC category was different.

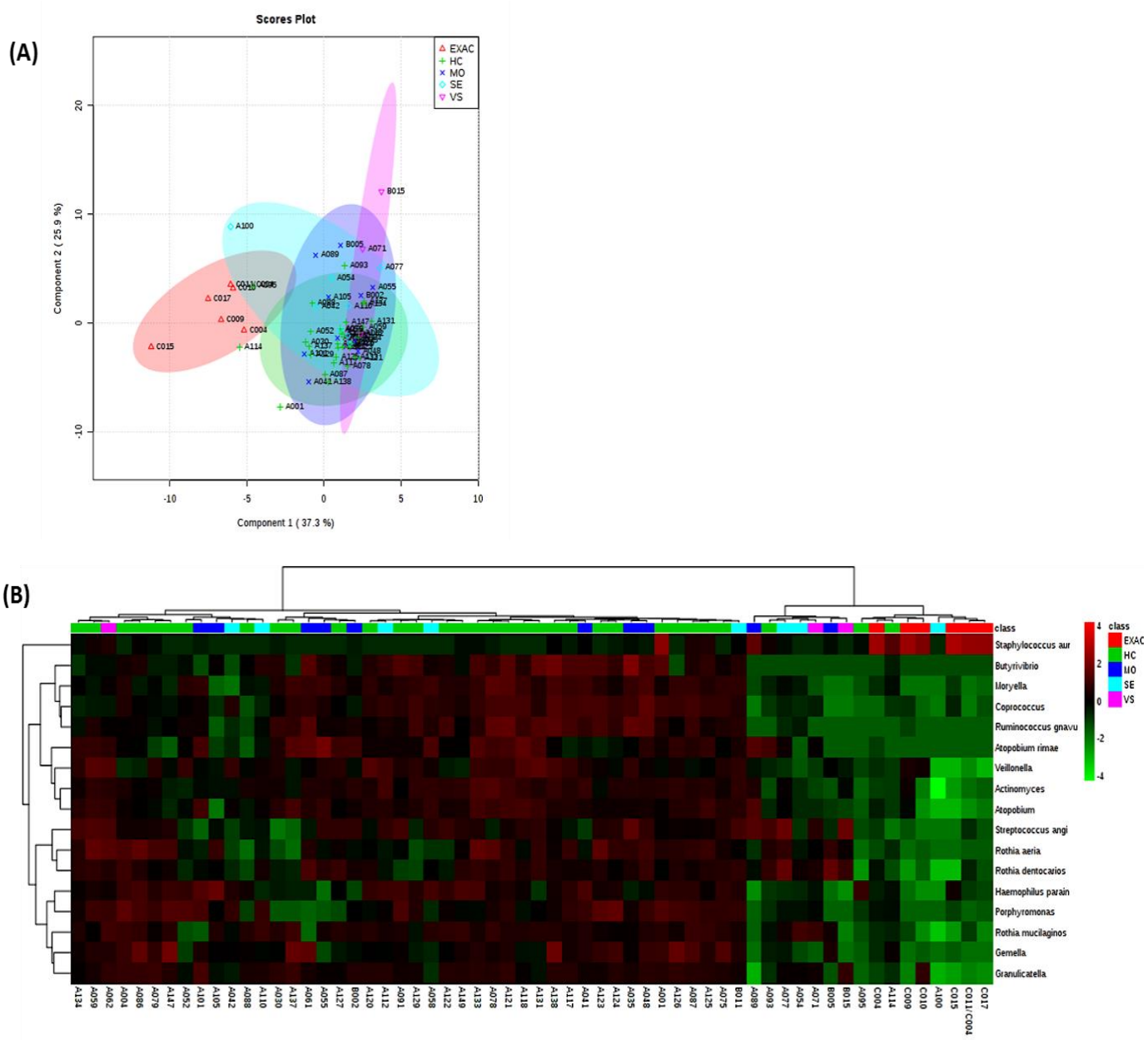


Figure 4.15 | Identifying genera which differed according to COPD classification MILD (MI), MODERATE (MOD), SEVERE (SEV) and VERY SEVERE (VS) and Healthy controls (HC)
(A) PCA of genera abundance in each COPD classification **(B)** Plot of a heat map and dendrogram of genera showing significant (P <0.05) variation between the classes as defined by ANOVA and plotted for each sample.

These data were also plotted for mean values for each bacterial genus for each COPD category (Figure 4.16). This indicated three patterns of response to progressive COPD; an increase seen only in the EXAC category (Figure 4.16B) and seen only on *S. aureus;* progressive reduction

in abundance as COPD progresses (Figure 4.16C) and finally, those which exhibited a significant reduction only in the EXAC category (Figure 4.16D). Such very different types of response to progressive COPD would argue for specific responses to patient treatment regimens which include the use of antibiotics and/or inflammatory status.

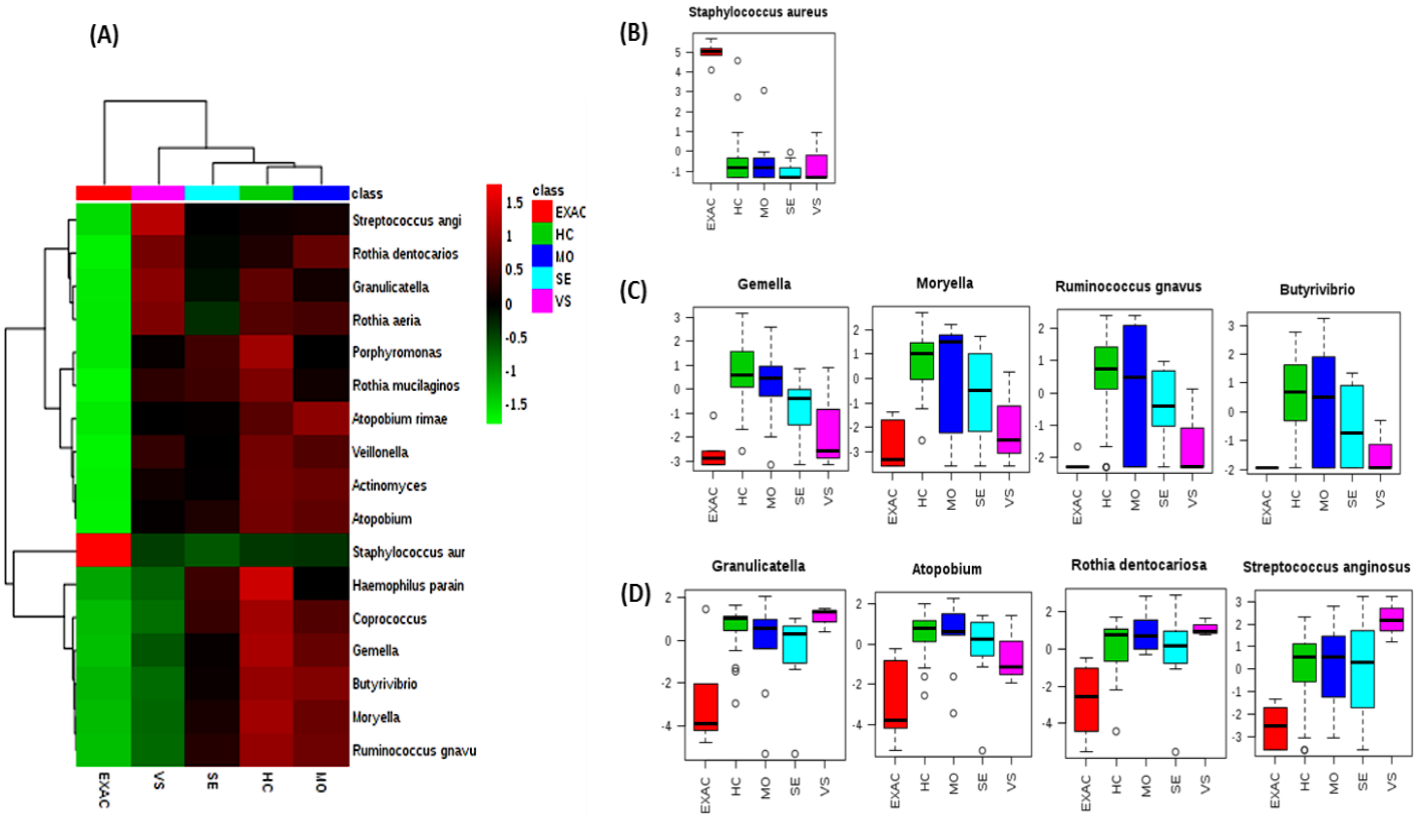


Figure 4.16 | Genera that showed significant (P <0.05) variation between the COPD experimental classes as defined by ANOVA

Results were plotted as **(A)** heat map of mean values and box and whisker of **(B)** genera showing increases in the COPD exacerbator class **(C)** general showing progressive decreases in abundance as COPD progresses and **(D)** genera showing a reduction in the EXAC class.

4.3.3 | Saliva Microbiome and Lung Cancer

Given the microbiomic changes that were occurring in the COPD stages (Figure 4.16), possible differences linked to LC were examined. Focusing on genera, PERMANOVA failed to demonstrate any significant differences occurring with lung cancer (Figure 4.17).

The genera with the LC and HC categories were plotted (Figure 4.18), but these did not exhibit any significant differences in α–diversity.

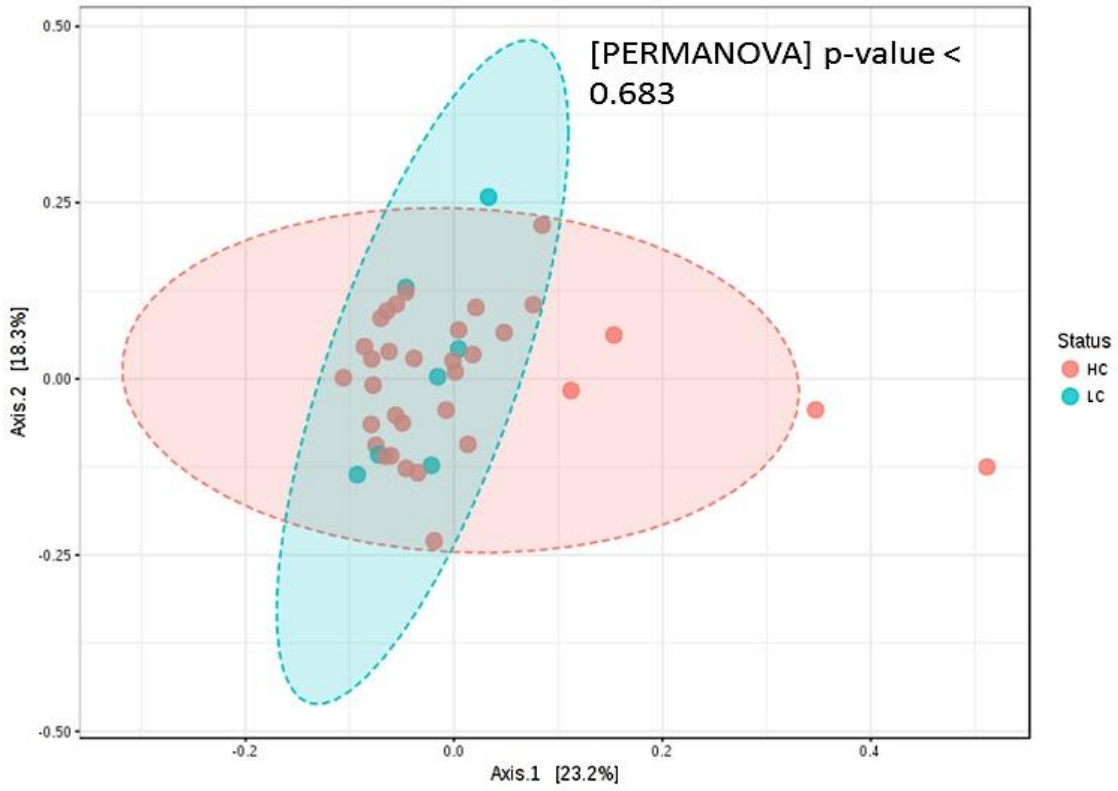


Figure 4.17 | Assessing variation in the microbiome genera linked to Lung Cancer (LC) and Healthy Controls (HC) displayed through Principal Coordinate analysis (PCoA) with estimates of significance derived by PERMANOVA

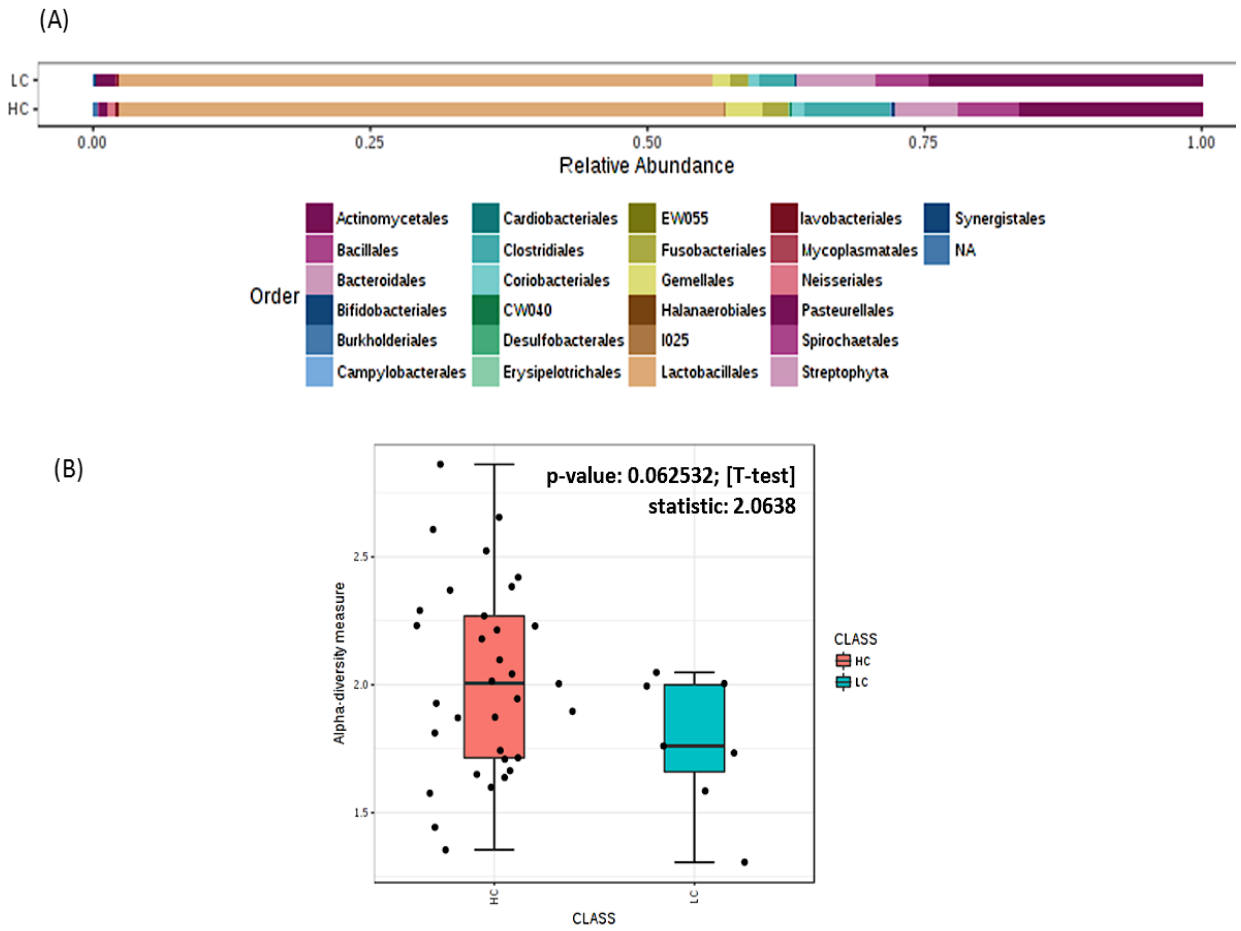


Figure 4.18 | Features in the microbiome genera linked to Lung Cancer (LC) and healthy controls (HC) displayed (A) bar graphs and (B) α-diversity

4.4 | Discussion

Due to the easiness in sampling and possible usefulness in disease biomarkers, human saliva has potential to be a valid biofluid for future microbiomic analyses. Previous studies have managed to demonstrate that human saliva may offer understanding on insights of cancer diseases such as oral, breast and pancreatic (Sugimoto *et al*., 2010), on obesity (Matias *et al*., 2012) and dental caries (Yang *et al.*, 2012).

In this segment of work, the salivary microbiome of 74 patients was sampled over a seven-month period. All samples underwent 16S rRNA amplicon sequencing (V3 – V4) to identify the taxonomic composition of their salivary microbiome. Additionally, the estimated salivary microbial load was measured on a group of 28 patients.

4.4.1 | Bacterial Load

The bacterial load of human saliva may have potential to be a marker of immunity and an indicator of bacterial levels throughout seasonal changes or severity of disease (Jones *et al*., 2014). Moreover, the salivary bacterial load has indeed been suggested to originate from other causes i.e. respiratory diseases than dental conditions as gingivitis and/or periodontal disease (Mantilla Gomez *et al*., 2001).

In this portion of work, human saliva was assessed as a potential biosample source for the 16S bacterial microbiome through the qPCR analysis. As displayed in Figure 4.5 and detailed in Table 4.2, saliva samples that contained higher levels of bacterial load belonged mostly to HC participants or patients that were not undergoing antimicrobial treatment. Variables such as smoking status and pack years did not seem to influence the outcome of the bacterial load analysis. Patients that were on antibiotic medication at time of recruitment displayed lower levels of bacterial load.

These results were in agreement with the findings of Erb-Downward *et al*. (2011) who used 16S pyrosequencing analysis to examine four COPD patients, seven healthy smokers and three non-smokers. The outcome did not indicate a significant differentiation between the lung microbiomes of the three groups. However, the lung microbiome diversity of the COPD patients was shown to be reduced in comparison to healthy smokers.

Similarly, the bacterial load analysis conducted on the HC, COPD, COPD Exacerbation and LC experimental classes (Figure 4.6) indicated patients which were undergoing an exacerbation had significantly lower loads when compared to the other categories. Therefore, such observations suggest that imposing stress on the lung via smoking or certain pathological situation, such as COPD exacerbations, results in a reduction in certain bacterial populations. This could arise from inflammatory or immunological responses in the lung and may be associated with reduced lung function under both situations. Equally, it could be that the altered lung function itself may influence the lung microbiome.

4.4.2 | Amplicon Sequencing

The initial analysis of the core microbiome has shown that *Streptococcus* was prevalent in every saliva sample (Figure 4.8). The permutational multivariate analysis of variance indicated the COPD Exacerbation class as the most distinctive (Figure 4.9). These differences were assessed at different levels of taxonomic classification: phyla, genus and species.

The bacterial genera and species displayed similar bacterial abundances between three experimental classes COPD, HC and LC, but significantly different abundances in the COPD Exacerbation group. The COPD Exacerbation class was associated with a loss in a range of bacterial genera. However, it displayed an increase in the *Staphylococcus* genus (Figures 4.11 and 4.13), particularly the species *S. aureus* (Figure 4.14, Table 4.3).

Interestingly, *S. aureus* was by far the most abundant out of the *Staphylococcus* species in all saliva samples (Table 4.3). Increases in *S. aureus* may reflect interactions with other species such as the fungus *Candida albicans*, in forming multispecies biofilms with it to help penetrate human tissue (Schlecht *et al.*, 2015). The loss in bacterial diversity is a major feature of the COPD exacerbation class, posing the question of why this should be the case.

The relative abundance analysis of LC and HC genera (Figure 4.18) has not exhibited great differences between the two groups and the α–diversity measurement displayed a reduction in the LC group in comparison with HC class.

Considering the data together, there is the possibility that the reduction in the numbers of certain genera cannot be linked to any abilities to act as an opportunistic pathogen or respond to host immune status. Instead, it may reflect the outcomes of patients being treated with antibiotics. This would also agree with the reduction in overall bacterial load seen in the COPD Exacerbator group. Given that *S. aureus* is breaking this trend it may be that antibiotic treatments could be selecting for resistant bacterial strains of a significant pathogen in the oral cavity. On the other hand, the alteration in microbial community content during exacerbation episodes may be an adaptation in trying to protect the lung from respiratory viral infections.

A more in-depth perception of pathogenic bacteria in relation to respiratory diseases and healthy individuals was needed in this segment of work. This in turn could be used to widen the understanding of bacterial variability in humans and establish whether these findings are common to respiratory diseases in general, or solely under-disease conditions.

In depth examinations of the human microbiome in both healthy individuals and patients affected by COPD or lung cancer may reveal new biomarkers (Spratlin, Serkova, Eckhardt, 2009). These could help in pulmonary diagnosis and monitoring as well as treatment

development. In this context, Huang *et al*. (2010) used a 16S rRNA PhyloChip on a group of eight patients suffering of COPD, with the aim of identifying significant pathogenic bacteria. They found 75 bacterial taxa that were detected in all COPD study participants, even in the patients who were undergoing antibiotic treatment. These results suggest that detection of pathogenic bacteria can potentially be used in identifying therapeutic targets to improve exacerbation treatment for individuals suffering of COPD and potentially be applied to patients suffering of lung cancer or other respiratory diseases.

4.5 | Conclusions and Future Work

The findings suggest that from a small sample size, through microbiomic sequencing analyses, the taxonomic composition of bacterial population can be defined. Based on this, lung function with respect to specific respiratory diseases can be suggested. The conventional notion which states that lungs are a sterile environment has long been contradicted (Charlson *et al*., 2011). Notably, the lung microbiota has been analysed in numerous studies in patients suffering from Cystic Fibrosis and asthmatic patients. However, new investigations that focus on other respiratory diseases with wider impact such as COPD and lung cancer are required. These studies are needed in order to establish whether the patients' lung microbiome suffers alterations specifically caused by the disease symptoms or they are just a manifestation of the lung condition.

CHAPTER 5 | General Discussion and Conclusions

In this research project, it has been proven that changes in certain metabolites and bacteria can allow the differentiation between COPD, LC and healthy patients.

The research assignment addressed this by employing metabolomics fingerprinting on a sample cohort. This allowed the identification of a range of metabolites that were proficient enough to differentiate between certain respiratory diseases and healthy controls (never, current, ex-smokers). Although the identity of individual metabolites was not established, potential biomarkers were suggested, that allowed discrimination between the four groups created COPD, LC, HC and Other.

The metabolomics fingerprinting technique based on saliva samples, has the potential of generating a high-throughput, cost-effective, non-invasive method of screening and diagnosing patients suffering of specific diseases, not only of the pulmonary nature. A larger cohort would undeniably be needed in order to confirm the accuracy of the results presented in this study and ascertain they are not artefacts. However, the technique of flow infusion electrospray mass spectrometry in saliva metabolomics fingerprinting suggests a promising outlook with regards to disease types and stages.

The second approach of this research project was to employ 16S rRNA amplicon sequencing in order to establish the distinct bacterial taxonomic composition of COPD, COPD Exacerbation, HC and LC classes. Patients which were undergoing an exacerbation had significantly lower bacterial loads when compared to the other categories. This in turn may offer insights of the clinical features of respiratory diseases in relation to the levels of bacterial load during stable conditions or exacerbations.

As with the lung metabolome investigation, the lung microbiome would require a larger cohort as well, in order to accurately discriminate between COPD, lung cancer, healthy controls and other respiratory diseases down to a genus and possibly species level. Additionally, a complete set of variables such as FEV_1%, co-morbidities, stable or exacerbated condition, would be required for all patients for the completeness of results.

The metabolomics and microbiomic approaches presented in here, both show differentiation between the COPD, lung cancer, other respiratory disease and healthy controls. With regards to the metabolome and microbiome work, insufficient samples were available in order to conduct both profiling techniques. Therefore, the metabolomics and microbiomic link of every sample was not possible. However, even if these results have been conducted at a small scale, saliva offered an unexpected reliability in the differentiation process for COPD, lung cancer and healthy controls. In both metabolomics (PCA analysis between classes, group heatmap) as well as microbiomic analyses (PCoA, PCA genera abundance of COPD stages, group heatmap), COPD and COPD Exacerbation groups showed an evident differentiation compared to LC, HC, Other classes. In the metabolomics Random Forest Plot, the COPD group displayed upregulated metabolites suggesting inflammation and in the microbiomics *a*-biodiversity measurement the COPD exacerbator group displayed a significant reduction suggesting potential immunological responses linked to reduced lung function.

Metabolomic differentiation between the four classes was more evident in comparison with the microbiomic one. Specifically between LC and HC groups, but it must be mentioned that with the metabolomic analysis the PLS-DA was performed, which is a supervised analysis aiming to provide differentiation between classes.

Finally, one important future aim will be to understand the human microbiome through its metabolic products. The production of particular metabolites by bacteria in the human respiratory system, may provide insights with regards to the microbiomic variation.

6 | REFERENCES

AAS, J.A., PASTER, B.J., STOKES, L.N., OLSEN, I. and DWEHIRST, F.E., 2005. Defining the Normal Bacterial Flora of the Oral Cavity. *Journal of Clinical Microbiology*, 43(11), pp.5721-32.

ADAWY, A.E., 2016. Spirometry Basics. [online] Available at: http://www.slideshare.net/ashrafeladawy/spirometry-basics-2 [Accessed 6 Jan. 2017].

AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY, 2013. Environmental Health and Medicine Education. Radon Toxicity Is at Risk of Radon Exposure?. [online] Available at: https://www.atsdr.cdc.gov/csem/csem.asp?csem=8&po=7 [Accessed 2 Jan. 2017].

AHL, J., LITTORIN, N., FORSGREN, A., ODENHOLD, I., RESMAN, F AND RIESBECK, K., 2013. High incidence of septic shock caused by *Streptococcus pneumoniae* serotype 3: a retrospective epidemiological study. *BMC Infectious Diseases*, 13, pp.492.

ALAMDARI, D.H., MOBARHAN, M.G., TAVALLAIE, S., PARIZADEH, M.R., MOOHEBATI, M., GHAFOORI, F., BAJESTANI, S.M.R.K., PALETAS, K., PEGIOU, T., KOLIAKOS, G., 2008. Pro-oxidant-antioxidant balance as a new risk factor in patients with angiographically defined coronary artery disease. *Clin Biochem*, 41, pp.375-80.

ANDERSON, K.E., CARMELLA, S.G., YE, M., BLISS, R.L., LE, C., MURPHY, L. and HECHT, S.S., 2001. Metabolites of a Tabacco-Specific Lung Carcinogen in Nonsmoking Women Exposed to Environmental Tobacco Smoke. *JNCI*, 93, pp.378-381.

ASH, 2005. Smoking and Health Inequalities. [online] Available at: http://ash.org.uk/files/documents/ASH_82.pdf [Accessed 30 Dec. 2016].

ASTHMA UK, 2015. Symptoms of asthma. [online] Available at: https://www.asthma.org.uk/advice/understanding-asthma/symptoms/ [Accessed on 15 Jan. 2017].

ATKINSON, A.J., 2001. NCI-FDA Biomarkers Definitions Working Group; Biomarkers and endpoints: preferred definitions and conceptual framework. Clin. Pharmaco. Ther., 69, pp.89-95.

BALDWIN, D.R., DUFFY, S.W., WALD, N.J., PAGE, R., HANSELL, D.M. and FIELD, J.K., 2011. UK Lung Screen (UKLS) Nodule Management Protocol: Modelling of a Single Screen Randomised Controlled Trial of Low-Dose CT Screening for Lung Cancer. *Thorax*, 66(4), pp.308-313.

BANDHAKAVI, S., STONE, M.D., ONSONGO, G., VAN RIPER, S.K., Griffin, T.J., 2009. A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva. *J Proteome Res*, 8(12), pp.5590-600.

BARTSCH, H., NAIR, U., RISCH, A., ROJAS, M., WIKMAN, H., ALEXANDROV, K., 2000. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancer. *Cancer Epidemiol Biomarkers Prev*, 9(1), pp.3-28.

BASSIS, C.M., ERB-DOWNWARD, J.R., DICKSON, R.P., FREEMAN, C.M., SCHMIDT, T.M., YOUNG, V.B., CURTIS, J.L., HUFFNAGLE, G.B. 2015. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio*, 6, pp.e00037.

BATEMAN, E.D., HURD, S.S., BARNES, P.J., BOUSQUET, J., DRAZEN, J.M., FITZGERALD, M., GIBSON, P., OTHA, K., O'BYRNE, P.O., PEDERSEN, S.E., PIZZICHINI, E., SULLIVAN, S.D., WENZEL, S.E., ZAR, H.J., 2008. Global strategy for asthma management and prevention: GINA executive summary. *European Respiratory Journal*, 31, pp.143-178.

BAULD, L., 2006. Smoking treatment and inequalities in health. Presentation at UK *National Smoking Cessation Conference (UKNSCC)*.

BIBBY, B.G., 1949. What about saliva?. *Oral Surg Oral Med Pral Pathol*, 2, pp.72-81.

BIO-RAD LABORATORIES, INC., 2006. Real-Time PCR Application Guide. [online] Available at: http://www.gene-quantification.com/real-time-pcr-guide-bio-rad.pdf [Accessed 11 Sept. 2017].

BLANC, P., TOREN, K., 2007. Occupation in chronic obstructive pulmonary disease and chronic bronchitis: an update. *Int J Tuberc Lung Dis*, 11(3), pp.251-257.

BLASI, F., MANTERO, M., PIERACHILLE, S. AND TARSIA, P., 2012. Understanding the burden of pneumococcal disease in adults. *Clinical Microbiology and Infections*, 18(5), pp.7-14.

BODE, A.M., DONG, Z. and WANG, H., 2015. Cancer prevention and control: alarming challenges in China. *Natl Sci Rev*, 00, pp.1-11.

BOFFETTA, P., PERSHAGEN, G., JOCKEL, K.H., FORASTIERE, F., GABORIEAU, V., HEINRICH, J., JAHN, I., KREUZER, M., MERLETTI, F., NYBERG, F., ROSCH, F., SIMONATO, L., 1999. Cigar and pipe smoking and lung cancer risk: a multicentre study from Europe. J Natl Cancer Inst, 91(8), pp.697-701.

BONNANS, C., CHOU, J., & WERB, Z., 2014. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology*, 15, pp.786-801.

BONNE, N.J. AND WONG, D.T.W., 2012. Salivary biomarker development using genomic, proteomic and metabolomics approaches. *Genome Medicine*, 4, pp.82

BORON, W.F., 2003. Medical Physiology: A Cellular and Molecular Approach. *Elsevier/Saunders*, pp.928.

BRITISH LUNG FOUNDATION, 2012. Asthma statistics. [online] Available at: https://statistics.blf.org.uk/asthma [Accessed 15 Jan. 2017].

BRITISH LUNG FOUNDATION, 2012. Bronchiectasis. [online] Available at: https://www.blf.org.uk/support-for-you/bronchiectasis/what-is-it [Accessed 18 Jan. 2017].

BRITISH LUNG FOUNDATION, 2012. Chronic obstructive pulmonary disease (COPD) statistics. [online] Available at: https://statistics.blf.org.uk/copd [Accessed 5 Jan. 2017].

BRITISH LUNG FOUNDATION, 2012. Idiopathic pulmonary fibrosis (IPF). [online] Available at: https://www.blf.org.uk/support-for-you/idiopathic-pulmonary-fibrosis-ipf [Accessed 18 Jan. 2017].

BRITISH LUNG FOUNDATION, 2012. Pneumonia statistics. [online] Available at: https://statistics.blf.org.uk/pneumonia [Accessed 11 Jan. 2017].

BRITISH LUNG FOUNDATION, 2016. Lung disease in the UK – big picture statistics. [online] Available at: https://statistics.blf.org.uk/lung-disease-uk-big-picture [Accessed 25 July 2016].

BRITISH THORACIC SOCIETY, 2006. Burden of Lung Disease. [online] Available at: https://www.brit-thoracic.org.uk/delivery-of-respiratory-care/burden-of-lung-disease/ [Accessed 22 July 2016].

BROWN, J.S., 2012. Community-acquired pneumonia. Clinical Medicine, 12(6), pp.538-543.

BYRNE, M.M., WEISSFELD, J. and ROBERTS, M.S., 2008 Anxiety, Fear of Cancer, and Perceived Risk of Cancer following Lung Cancer Screening. *Medical Decision Making*, 28(6), pp.917-925.

CAMANOCHA, A. and DEWHIRST, F.E., 2014. Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions. *J Oral Microbiol*, 6, pp.10.3402/jom.v6.25468.

CAMERON, S.J., 2015. Charting Human Microbiome and Metabolome Changes in Disease and Stress. Ph.D. *Aberystwyth University - IBERS*, pp.1-226.

CAMERON, S.J., HUWS, S.A., HEGARTY, M.J., SMITH, D.P., MUR, L.A., 2015. The human salivary microbiome exhibits temporal stability in bacterial diversity. *FEMS Microbiol Ecol*, 91(9), pp.091.

CAMERON, S.J.S., LEWIS, K.E., BECKMANN, M., ALLISON, G.G., GHOSAL, R., LEWIS, P.D., MUR, L.A.J., 2016. The metabolomics detection of lung cancer biomarkers in sputum. *Lung Cancer*, 94, pp.88-95.

CANCER RESEARCH UK, 2014. Cancer mortality statistics. [online] Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality#heading-One [Accessed 12 Dec. 2016].

CANCER RESEARCH UK, 2016. Cancer and health inequalities: An introduction to current evidence. [online] Available at: http://www.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@pol/documents/generallcontent/crukmig_1000ast-3344.pdf [Accessed 30 Dec. 2016].

CAT COPD Assessment Test, 2012. Expert guidance on frequently asked questions. Healthcare Professional Used Guide. [online] Available at: http://www.catestonline.org/images/UserGuides/CATHCPUser%20guideEn.pdf [Accessed 7 Jan. 2017].

CHAKRABORTY, D., MONDAL, N.K., DATTA, J.K., 2014. Indoor pollution from solid biomass fuel and rural health damage: A micro-environmental study in rural area of Burdwan, West Bengal. *International Journal of Sustainable Built Environment*, 3(2), pp.262-271.

CHAKRAVORTY, S., HELB, D., BURDAY, M., CONNELL, N., ALLAND, D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*, 69(2), pp.330-339.

CHARLSON, E.S., BITTINGER, K., HAAS, A.R., FITZGERALD, A.S., FRANK, I., YADAV, A., BUSHMAN, F.D. and COLLMAN, R.G. 2011. Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract. *American Journal of Respiratory and Critical Care Medicine*, 184(8), pp.957-63.

CHEN, Q., DEEB, R.S., MA, Y., STAUDT, M., CRYSTAL, R.G., GROSS, S.S., 2015. Serum Metabolite Biomarkers Discriminate Healthy Smokers from COPD Smokers. *PLoS ONE*, 10(12), pp.e0143937.

CHEN, Y., DALES, R., TANG, M., KREWSKI, D., 2002. Obesity May Increase the Incidence of Asthma in Women but Not in Men: Longitudinal Observations from the Canadian National Population Health Surveys. *American Journal of Epidemiology*, 155(3), pp.191-197.

CILLONIZ, C., EWIG, S., MENENDEZ, R., FERRER, M., POLVERINO, E., REYES, S., GABARRUS, A., MARCOS, M.A., CORDOBA, J., MENSA, J., TORRES, A., 2012. Bacterial co-infection with H1N1 infection in patients admitted with community acquired pneumonia. *J Infect*, 65(3), pp.223-30.

COLLER, H.A., 2014. Is Cancer a Metabolic Disease?. *The American Journal of Pathology*, 184(1), pp.4-17.

COPES, R., SCOTT, J., 2007. Radon exposure: can we make a difference?. *Canadian Medical Association Journal*, 177(10), pp.1229-1231.

CORNFIELD, J., HAENSZEL, W., HAMMOND, E., LILIENFELD, A., SHIMKIN, M. and WYNDER, E., 2009. Smoking and lung cancer: recent evidence and a discussion of some questions. *International Journal of Epidemiology*, 38(5), pp.1175-1191.

CORRALES-MEDINA, V.F. and MUSHER, D.M., 2013. Immunomodulatory agents in the treatment of community-acquired pneumonia: a systematic review. *Journal of Infection*, 63(3), pp.187-199.
CRAIG, V.J., POLVERINO, F., LAUCHO-CONTRERAS, M.E., SHI, Y., LIU, Y., OSORIO, J.C., TESFAIGZI, Y., PINTO-PLATA, V., GOCHUICO, B.R., ROSAS, I.O., OWEN, C.A., 2014. Mononuclear Phagocytes and Airway Epithelial Cells: Novel Sources of matrix Metalloproteinase-8 (MMP-8) in Patients with Idiopathic Pulmonary Fibrosis. *PloS One*, 9, pp.e97485.
DALLAS, M.E., 2013. Pneumonia Symptoms and Causes. Everyday Health. [online] Available at: http://www.everydayhealth.com/pneumonia/guide/symptoms/ [Accessed 11 Jan. 2017].
DAWES, C., 1969. The effects of flow rate and duration of stimulation on the concentrations of protein and the main electrolytes in human parotid saliva. *Arch Oral Biol*, 14, pp.277-94.

DEEPE Jr, G.S. 2018. Outbreaks of histoplasmosis: The spores set sail. *PLoS Pathog*, 14(9), pp.e1007213.

DEJA, S., KOWAL, A., BARG, W., STANIMIROVA, I., KORZENIEWSKA, A., MLYNARZ, P., POREBSKA, I., ZABEK, A., PAWELCZYK, K., DASZYKOWSKI, M., JANKOWSKA, R., 2014. Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. *Journal of Pharmaceutical and Biomedical Analysis*, 100C, pp.369-380.

DICKSON, R.P., ERB-DOWNWARD, J.R., MARTINEZ, F.J., HUFFNAGLE, G.B. 2016. The Microbiome and the Respiratory Tract. *Annu Rev Physiol*, 78, pp.481-504.

DISEASES AND CONDITIONS, 2016. Heart diseases. [online] Available at: http://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease/art-20046167 [Accessed 11 Jan. 2017].

DOEHNER, W., VON HAEHLING, S. ANKER, S.D., LAINSCAK, M., 2009. Neurohormonal activation and inflammation in chronic cardiopulmonary disease: a brief systematic review. *Wien Klin Wochenschr*, 121, pp.293-6.

DONALDSON, G.C. AND WEDZICHA, J.A., 2014. The causes and consequences of seasonal variation in COPD exacerbations. *Int J Chron Obstruct Pulmon Dis*, 9, pp.1101-1110.

DONALDSON, G.C., SEEMUNGAL, T.A., BHOWMIK, A., WEDZICHA, J.A., 2002. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax*, 63(8), pp.753.

DRAPER, J., LLOYD, A.J., GOODACRE, R., BECKMANN, M., 2013. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics*, 9, pp.S4-S29.

DUCHEMANN, B., TRIBA, M.N., GUEZ, D., RZEZNIK, M., SAVARIN, p., NUNES, H., VALEYRE, D., BERNAUDIN, J.F., LE MOYEC, L., 2016. Nuclear magnetic resonance spectroscopic analysis of salivary metabolome in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis*, 33(1), pp.10-6.

DURHAM, A.L. and ADCOCK, I.M. 2015. The relationship between COPD and lung cancer. *Lung Cancer*, 90(2), pp.121-127.

EMWAS, A.H., 2015. The strengths and weaknesses of NMR spectroscopy and mass spectroscopy with particular focus on metabolomics research. *Methods Mol Biol*, 1277, pp.161-93.

ENVIRONMENTAL PROTECTION AGENCY, 2017. Epidemiology & Health Effects of Cyanobacteria. [online] Available at: https://www.epa.gov/water-research/epidemiology-health-effects-cyanobacteria [Accessed 1 Dec. 2017].

EPSTEIN, J.B., GORSKY, M., GUGLIETTA, A., LE, N., SONIS, S.T., 2000. The correlation between epidermal growth factor levels in saliva and the severity of oral mucositis during oropharyngeal radiation therapy. *Cancer*, 89(11), pp.2258-65.

ESSENTIALS OF HUMAN PHYSIOLOGY, n.d. Composition of Saliva. [online] Available at: http://humanphysiology.tuars.com/program/section6/6ch4/s6ch4_6.htm [Accessed 30 Aug. 2016].

EUZEBY, J.P., 2013. Chloroflexi. List of Prokaryotic names with Standing in Nomenclature (LPSN). [online] Available at: https://web.archive.org/web/20130616035835/http://www.bacterio.cict.fr/classifphyla.html [Accessed 30 Nov. 2017].

FAHIM, A., CROOKS, M. AND HART, S.P., 2011. Gastroesophageal Reflux and Idiopathic pulmonary fibrosis: A Review. *Pulmonary Medicine*, 634613, pp.7.

FARNAUD, S.J.C., KOSTI, O., GETTING, S.J., RENSHAW, D., 2010. Saliva: Physiology and Diagnostic Potential in Health and Disease. *The Scientific World Journal*, 10, pp.434-456.

FAUCI, A.S. AND MORENS, D.M., 2012. The perpetual challenge of infectious diseases. *The New England Journal of Medicine*, 366(5), pp. 454-461.

FELDMAN, C. AND ANDERSON, R., 2009. New insights into pneumococcal disease. *Respirology*, 14(2), pp.167-179.

FELDMAN, C., 2007. Prognostic scoring systems: which one is best?. *Current Opinion in Infectious Diseases*, 20(2), pp.165-169.

FERREL, J.J., ZHANG, L., ZHOU, H., CHIA, D., ELASHOFF, D., AKIN, D., Paster, B.J., JOSHIPURA, K., WONG, D.T., 2012. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. *Gut*, 61, pp.582-588.

FIELD, R.W., STECK, D.J., SMITH, B.J., BRUS, C.P., FISHER, E.L., NEUBERGER, J.S., PLATZ, C.E., ROBINSON, R.A., WOOLSON, R.F. and LYNCH, C.F., 2000. Residential Radon Gas Exposure and Lung Cancer – The Iowa Radon Lung Cancer Study. *American Journal of Epidemiology*, 151(11), pp.1091-1102.

FILE JR., T.M. and MARRIE, T.J., 2010. Burden of community-acquired pneumonia in North American adults. *Postgraduate Medicine*, 122(2), pp.130-141.

FLYNN, M., BAKER, E.S., AND KASS, D.J., 2015. Idiopathic pulmonary fibrosis biomarkers: clinical utility and a way of understanding disease pathogenesis. *Dove Medical Press*, 2015(5), pp.21-33.

FORSBERG, L., LARSSON, C., SOFIADIS, A., LEWENSOHN, R., HOOG, A., LEHTIO, J., 2006. Pre-fractionation of archival frozen tumours for proteomics applications. *J Biotechnol*, 126(4), pp.582-4.

FOUNDATION FOR THE NATIONAL INSTITUTES OF HEALTH, 2011. The Biomarkers Consortium. [online] Available at: http://www.biomarkersconsortium.org/ [Accessed 23 Aug. 2016].

FULLER, D., MCGEER, A. AND LOW, D.E., 2005. Drug-resistant pneumococcal pneumonia: clinical relevance and approach to management. *European Journal of Clinical Microbiology and Infectious Diseases*, 24(12), pp.780-788.

GALIGIANI, J.N. 1993. Coccidiodomycosis. *West J Med*, 159(2), pp.153-171.

GARCHA, D.S., THURSTON, S.J, PATEL, A.R.C., MACKAY, A.J., GOLDRING, J.J.P., DONALDSON, G.C., MCHUGH, T.D. AND WEDZICHA, J.A., 2012. Changes in Prevalence and Load of Airway Bacteria Using Quantitative PCR in Stable and Exacerbated COPD. *Thorax*, 67(12), pp.1075-80.

GENE-QUANTIFICATION, 2013. Transcriptional Biomarkers. [online] Available at: http://transcriptional-biomarkers.gene-quantification.info/ [Accessed 22 Aug. 2016].

GENERAL HOUSEHOLD SURVEY, 2005. Office for National Statistics and Vital Statistics Division. [online] Available at: https://discover.ukdataservice.ac.uk/catalogue/?sn=5640 [Accessed 30 Dec. 2016].

GLOBAL INITIATIVE FOR CHRONIC OBSTRUCTIVE LUNG DISEASE (GOLD), 2013. Global Strategy for The Diagnosis, Management, And Prevention of Chronic Obstructive Pulmonary Disease. [online] Available at: http://goldcopd.org [Accessed 7 Jan. 2017].

GLOBAL INITIATIVE FOR CHRONIC OBSTRUCTIVE LUNG DISEASE (GOLD), 2014. Global Strategy for the Diagnosis, management and Prevention of COPD. [online] Available at: http://goldcopd.org [Accessed 5 Jan. 2017].

GREENING, A.P., IND, P.W., NORTHFIELD, M., SHAW, G., 1994. Added salmeterol versus higher-dose corticosteroid in asthma patients with symptoms on existing inhaled corticosteroid. *The Lancet*, 344(8917), pp.219-224.

HALBERT, R.J., NATOLI, J.L., GANO, A., BADAMGARAV, E., BUIST, A.S. and MANNINO, D.M., 2006. Global Burden of COPD: Systematic Review and Meta-Analysis. *The European Respiratory Journal*, 28(3), pp.523-32.

HAN, M.K., HUANG, Y.J., LIPUMA, J.J., BOUSHEY, H.A., BOUCHER, R.C., COOKSON, W.O., CURTIS, J.L., ERB-DOWNWARD, J., LYNCH, S.V., SETHI, S., TOEWS, G.B., YOUNG, V.B., WOLFGANG, M.C., HUFFNAGLE, G.B. and MARTINEZ, F.J., 2012. Significance of the Microbiome in Obstructive Lung Disease. *Thorax*, 67(5), pp.456-63.

HAN, Y.W., 2015. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Curr Opin Microbiol*, 0, pp.141-147.

HANAHAN, D., WEINBERG, R.A., 2000. The hallmarks of cancer. *Cell*, 100(1), pp.57-70.

HANANIA, N.A., MANNINO, D.M., YAWN, B.P., MAPEL, D.W., MARTINEZ, F.J., DONOHUE, J.F., KOSINSKI, M., RENDAS-BAUM, R., MINTZ, M., SAMUELS, S., JHINGRAN, P., DALAL, A.A., 2010. Predicting risk of airflow obstruction in primary care: Validation of the lung function questionnaire (LFQ). *Respir Med*, 104(8), pp.1160-70.

HEALTH AND SAFETY EXECUTIVE, 2016. Work-Related Chronic Obstructive Pulmonary Disease in Great Britain 2016. [online] Available at: http://www.hse.gov.uk/statistics/causdis/copd/ [Accessed 4 Jan. 2017].

HEALTHLINE, 2015. Non-Small Cell Lung Cancer vs Small Cell: Types, Stages, Symptoms, and Treatment. [online] Available at: http://www.healthline.com/health/lung-cancer/non-small-cell-lung-cancer-vs-small-cell#2 [Accessed 2 Jan. 2017].

HEALTHWISE, 2015. Salivary Gland Scan. [online] Available at: https://myhealth.alberta.ca/ [Accessed 18 May 2016].

HECHT, S.S., 1999. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst*, 91(14), pp.1194-210.

HECHT, S.S., 2000. Inhibition of carcinogenesis by isothiocyanates. Drug. *Metabol Rev*, 32, pp.395-411.

HECHT, S.S., 2002. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *The Lancet Oncology*, 3(8), pp.461-469.

HECHT, S.S., HOCHALTER, J.B., VILLATA, P.W., MURPHY, S.E., 2000. 2'-Hydroxylation of nicotine by cytochrome P450 2A6 and human liver microsomes: formation of a lung carcinogen precursor. *Proc Natl Acad Sci USA*, 97(23), pp.12493-7.

HECHT, S.S., HOFFMANN, D., 2000. Re: Cigar Smoking in Men and Risk of Death from Tobacco-Related Cancers. *Journal of the National Cancer Institute*, 92(24), pp.2040-2040.

HECHT, S.S., KENNEY, P.M.J., WANG, M., UPADHYAYA, P., 2000. Dose-response study of myo-inositol as an inhibitor of lung tumorigenesis induced in A/J mice by benzo[a]pyrene and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane. *Cancer Lett*, 167, pp.1-6.

HECHT, S.S., Ye, M., CARMELLA, S.G., FREDRICKSON, A., ADGATE, J.L., GREAVES, I.A, CHURCH, T.R, RYAN, A.D., MONGIN, S.J., SEXTON, K., 2001. Metabolites of a tobacco-specific lung carcinogen in the urine of elementary school-aged children. *Cancer Epidemiol Biomarkers Prev*, 10(11), pp.1109-16.

HIRSCHMANN, J.V., 2000. Do Bacteria Cause Exacerbations of COPD? *Chest*, 118(1), pp.193-203.

HO, P.L., QUE, T.L., NG, T.K., CHIU, S.S., YUNG, R.W. AND TSANG, K.W., 2006. Clinical outcomes of bacteremic pneumococcal infections in an area with high resistance. *European Journal of Clinical Microbiology & Infectious Diseases*, 25(5), pp.323-327.

HOFFMAN, D., HOFFMAN, I., EL-BAYOUMY, K., 2001. The less harmful cigarette: a controversial issue. A tribute to Ernst L. Wynder. *Chem Res Toxicol*, 14(7), pp.767-90.

HOUGHTON, A. M., 2013. Mechanistic links between COPD and lung cancer. *Nature Reviews Cancer*, 13, pp.233-245.

HUANG, Y.J., KIM, E., COX, M.J., BRODIE, E.L., BROWN, R., WEINER-KRONISH, J.P. and LYNCH, S.V., 2010. A Persistent and Diverse Airway Microbiota Present During Chronic Obstructive Pulmonary Disease Exacerbations. *OMICS: A Journal of Integrative Biology*, 14(1), pp.9-59.

HULKA, B.S., 1990. Overview of biological markers. In: Biological markers in epidemiology. New York: Oxford University Press, pp.3-15.

HUYGHE, A., FRANCOIS, P., CHARBONNIER, Y., TANGOMO-BENTO, M., BONETTI, E.J., PASTER, B.J., BOLIVAR, I., BARATTI-MAYER, D., PITTET, D., SCHRENZEL, J., Novel microarray design strategy to study complex bacterial communities. *Appl Environ Microbiol*, 74, pp.1876-1885.

INFANTE, M., CAVUTO, S., LUTMAN, F.R., BRAMBILLA, G., CHIESA, G., CERESOLI, G., PASSERA, E., ANGELI, E., CHIARENZA, M., ARANZULLA, G., CARIBONI, U., ERRICO, V., INZIRILLO, F., BOTTONI, E., VOULAZ, E., ALLOISIO, M., DESTRO, A., RONCALLI, M., SANTORO, A., RAVASI, G. and DANTE STUDY, G., 2009. A Randomized Study of Lung Cancer Screening with Spiral Computed Tomography – Three-Year Results from the Dante Trial. *American Journal of Respiratory and Critical Care Medicine*, 180(5), pp.445-453.

IORGULESCU, G., 2009. Saliva between normal and pathological. Important factors in determining systemic and oral health. *J Med Life*, 2, pp.303-7.

JARVIS, M., 2001. The challenge for reducing inequalities: analysis of the General Household Survey 1998. Presentation to a *Department of Health* seminar.

JEMAL, A., CENTER, M.M, DESANTIS, C. and WARD, E.M., 2010. Global Patterns of Cancer Incidence and Mortality Rates and Trends. *Cancer Epidemiology Biomarkers and Prevention*, 19(8), pp.1893-1907.

JENKINSON, H.F. and LAMONT, R.J., 2005. Oral Microbial Communities in Sickness and in Health. *Trends in Microbiology*, 13(12), pp.589-95.

JONES, A.W., CAMERON, S.J.S., THATCHER, R., BEECROFT, M.S., MUR, L.A.J. and DAVISON, G., 2014. Effects of Bovine Colostrum Supplementation on Upper Respiratory Illness in Active Males. *Brain, Behaviour, and Immunity*, 39, pp.194-203.

JONES, P.W., 2009. Health Status and the Spiral of Decline. *COPD*, 6(1), pp.59-63.

JONES, P.W., HARDING, G., BERRY, WIKLUND, I., CHEN, W-H., KLINE LEIDY, N., 2009. Development and first validation of the COPD Assessment Test. *European Respiratory Journal*, 34, pp.648-654.

KANVAH, S., JOSEPH, J., SCHUSTER, G., BARNETT, R., CLEVELAND, C. and LANDMAN, U., 2010. Oxidation of DNA: Damage to Nucleobases. *Accounts of Chemical Research*, 43(2), pp.280-287.

KATAJAMA, M. AND ORESIC, M., 2005. Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics, 6, pp.179.

KHALIL, N., CHURG, A., MULLER, N., O'CONNOR, R., 2007. Environmental, Inhaled and Ingested Causes of Pulmonary Fibrosis. *Toxicologic Pathology*, 35(1), pp.86-96.

KLINDWORTH, A., PRUESSE, E., SCHWEER, T., PEPLIES, J., QUAST, C., HORN, M. and GLOCKNER, F.O., 2013. Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies. *Nucleic Acids Research*, 41(1), p.e1.

KOLDITZ, M., EWIG, S. AND HOFFKEN, G., 2013. Management-based risk prediction in community-acquired pneumonia by scores and biomarkers. *European Respiratory Journal*, 41(4), pp.974-984.

KRIEG, N.R., LUDWIG, W., WHITMAN, W.B., HEDLUND, B.P., PASTER, B.J., STALEY, J.T., WARD, N., BROWN, D., PARTE, A., 2010. *Bergey's Manual of Systematic Bacteriology,* 2nd ed., New York: Springer, p.908. British Library no. GBA561951.

KUMADA, Y., ZHAO, C., ISHIMURA, R., IMANAKA, H., IMAMURA, K., NAKANISHI, K., 2007. Protein-protein interaction analysis using an affinity peptide tag and hydrophilic polystyrene plate. *J Biotechnol*, 128(2), pp.354-61.

LABROOTS, Marker, K., 2015. Finally Catching Small Cell Lung Cancer Before Metastasis. [online] Available at: https://www.labroots.com/trending/immunology/1941/finally-catching-small-cell-lung-cancer-before-metastasis [Accessed 3 Jan. 2017].

LANG, B.R., 1991. *Chlamydia pneumonia* as a differential diagnosis? Follow-up to a case report on progressive pneumonitis in an adolescent. *Patient Care*.

LEE, Y.-H., WONG, D.T., 2009. Saliva: An emerging biofluid for early detection of disease. *Am J Dent*, 22(4), pp.241-248.

LEVINSON, M.E., 2001. Pneumonia, including necrotizing pulmonary infections (lung abscess). *HARRISONS PRINCIPLES OF INTERNAL MEDICINE*, 2, pp.1475-1484.

LIANG, G., LIU, H., LI, X. AND ZHANG, A.-H., 2016. High-throughput metabolomics analysis discovers salivary biomarkers for predicting mild cognitive impairment and Alzheimer's disease. *RSC Advances*, 6, pp.75499-75504.

LIN, P.-H., YEH, S.-K., HUANG, W.-C., CHEN, H.-Y., CHEN, C.-H., SHEU, J.-R., LIN, C.-T., HUANG, Y.-K., 2013. Research performance of biomarkers from biofluids in periodontal disease publications. *Journal of Dental Sciences*, 10, pp.61-67.

LIPPI, G., MESCHI, T. AND CERVELLIN, G., 2011. Inflammatory biomarkers for the diagnosis, monitoring and follow-up of community-acquired pneumonia: clinical evidence and perspectives. *European Journal of Internal Medicine*, 22(5), pp.460-465.

LIU, B., FALLER, L.L., KLITGORD, N., MAZUMDAR, V., GHODSI, M., SOMMER, D.D., GIBBONS, T.R., TREANGEN, T.J., CHANG, Y.-C., LI, S., STINE, O.C., HASTURK, H., KASIF, S., SEGRE, D., POP, M. and AMAR, S., 2012. Deep Sequencing of the Oral Microbiome Reveals Signatures of Periodontal Disease. *PLoS ONE*, 7(6), p.e37919.

MACKENZIE, G., 2016. The definition and classification of pneumonia. *Pneumonia*, 8, pp.14.

MADIGAN, M. and MARTINKO, J., 2005. *Brock Biology of Microorganisms*, 11th ed., Prentice Hall.

MADSEN, R., LUNDSTEDT, T. AND TRYGG, J., 2010. Chemometrics in Metabolomics: A Review in Human Disease Diagnosis. *Analytica Chimica Acta*, 659(1-2), pp.23-33.

MAHAJAN, G.B., 2012. Antibacterial agents from Actinomyces – a review. *Frontiers in Bioscience*, 4, pp.240-53.

MANTILLA GOMEZ, S., DANSER, M.M., SIPOS, P.M., ROWSHANI, B., van der VELDEN, U. and van der WEIJDEN, G.A., 2001. Tongue Coating and Salivary Bacterial Counts in Healthy/Gingivitis Subjects and Periodontitis Patients. *Journal of Clinical Periodontology*, 28(10), pp.970-978.

MARRIE, T.J., LAU, C.Y., WHEELER, S.L., WONG, C.J. and FEAGAN, B.G., 2000. Predictors of Symptom Resolution in Patients with Community-Acquired Pneumonia. *Clinical Infectious Diseases*, 31(6), pp.1362-1367.

MATHERS, C.D. AND LONCAR, D., 2006. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, 3(11), pp.442.

MATIAS, I., GATTA-CHERIFI, B., TABARIN, A., CLARK, S., LESTE-LASSERRE, T., MARSICANO, G., PIAZZA, P.V. and COTA, D., 2012. Endocannabinoids Measurement in Human Saliva as Potential Biomarker of Obesity. *PLoS One*, 7(7), p.e42399.

MCKEOWN, S., RICHTER, A.G., O'KANE, C., MCAULEY, D.F., THICKETT, D.R., 2009. MMP expression and abnormal lung permeability are important determinants of outcome in IPF. *Eur Respir J*, 33, pp.77-84.

MEIJVIS, S.C.A., HARDEMAN, H., REMMELTS, H.F., HEIJLIGENBERG, R., RIJKERS, G.T., VAN VELZEN_BLAD, H., VOORN, G.P., VAN DE GARDE, E.M.W., ENDEMAN, H., GRUTTERS, J.C., BOS, W.J.W., BIESMA, D.H., 2011. Dexamethasone and length of hospital stay in patients with community-acquired pneumonia: a randomised double-blind, placebo-controlled trial. *Lancet*, 277, pp.2023-30.

MEITES, S., ROGOLS, S., 1971. Amylase isoenzymes, *CRC Crit Rev Clin Lab Sci*, 2, pp.103-38.

MEYER, F., PAARMANN, D., D'SOUZA, M., OLSON, R., GLASS, E.M., KUBAL, M., PACZIAN, T., RODRIGUEZ, A., STEVENS, R., WILKE, A., WILKENING, J. and EDWARDS, R.A., 2008. The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes. *BMC Bioinformatics*, 9(1), p.386.

MIGLANI, D.C., RAGHUPATHY, E., 1968. Research on saliva. A-critique. *J Indian Dent Assoc*, 40, pp.131-2.

MIKAMI, K., SUZUKI, M., KITAGAWA, H., KAWAKAMI, M., HIROTA, N., YAMAGUCHI, H., NARUMOTO, O., KICHIKAWA, Y., KAWAI, M., TASHIMO, H., ARAI, H., HORIUCHI, T., SAKAMOTO, Y., 2007. Efficacy of Corticosteroids in the Treatment of Community-Acquired Pneumonia Requiring Hospitalization. *Lung*, 185(5), pp.249-255.

MILETICH, I., 2010. Introduction to salivary glands: Structure, function and embryonic development. *Front Oral Biol*, 14, pp.1-20.

MILLARES, L., FERRARI, R., GALLEGO, M., GARCIA-NUNEZ, M., PEREZ-BROCAL, V., ESPASA, M., POMARES, X., MONTON, C., MOYA, A., MONSO, E. 2014. Bronchial microbiome of severe COPD patients colonised by *Pseudomonas aeruginosa*. *Eur J Clin Microbiol Infect Dis*, 33, pp.1101-1111.

MORRIS, A., BECK, J.M., SCHLOSS, P.D., CAMPBELL, T.B., CROTHERS, K., CURTIS, J.L., FLORES, S.C., FONTENOT, A.P., GHEDIN, E., HUANG, L., JABLONSKI, K., KLEERUP, E., LYNCH, S.V., SODERGREN, E., TWIGG, H., YOUNG, V.B., BASSIS, C.M., VENKATARAMAN, A., SCHMIDT, T.M., WEINSTOCK, G.M., LUNG HIVMP. 2013. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*, 187, pp.1067-1075.

NAGLER, R., BEN-IZHAK, O., COHEN-KAPLAN, V., SHAFAT, I., VLADOVSKY, I., AKRISH, S., ILAN, N., 2007. Heparanase up-regulation in tongue cancer: tissue and saliva analysis. *Cancer*, 110(12), pp.2732-9.

NALYSNYK, L., CID-RUZAFA, J., ROTELLA, P., ESSER, D., 2012. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *European Respiratory review*, 21, pp.355-361.

NAWA, T., NAKAGAWA, T., MIZOUE, T., KUSANO, S., CHONAN, T., FUKAI, S. and ENDO, K., 2012. Long-term prognosis of patients with Lung Cancer Detected on Low-Dose Chest Computed Tomography Screening. *Lung Cancer*, 75(2), pp.197-202.

NHS choices, 2014, Asbestosis. [online] Available at: http://www.nhs.uk/conditions/Asbestosis/Pages/Introduction.aspx [Accessed 2 Jan. 2017].

NHS inform, 2015, Salivary gland cancer. [online] Available at: http://www.nhsinform.co.uk/health-library/articles/c/cancer-of-the-salivary-gland/introduction/ [Accessed 17 May 2016].

NICHOLAS, B., DJUKANOVIC, R., 2009. Induced sputum: a window to lung pathology. *Biochemical Society Transactions*, 37(4), pp.868- 872.

NIEDERMAN, M.S., 2007. Recent advances in community-acquired pneumonia: inpatient and outpatient. *Chest*, 131(4), pp.1205-1215.

NISHIMURA, K., IZUMI, T., TSUKINO, M., OGA, T., 2002. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. *Chest*, 121, pp.1434-1440.

NYSTAD, W., MEYER, H.E., NAFSTAD, P., TVERDAL, A., ENGELAND, A., 2004. Body Mass Index in Relation to Adult Asthma among 135,000 Norwegian Men and Women. *American Journal of Epidemiology*, 160(10), pp.969-976.

O'DWYER, D.N., DICKSON, R.P., MOORE, B.B. 2016. The Lung Microbiome, Immunity and the Pathogenesis of Chronic Lung Disease. *J Immunol*, 196(12), pp.4839-4847.

O'SHEA, K., CAMERON, S.J.S., LEWIS, K.E., CHUAN, L., MUR, L.A.J., 2016. Metabolomics-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochimica et Biophysica Acta*, 1860, pp.2682-2687.

Office of Disease Prevention and Health Promotion, 2014. Respiratory Diseases. [online] Available at: https://www.healthypeople.gov/2020/topics-objectives/topic/respiratory-diseases [Accessed 08 Aug. 2016].

OGSTON, A., 1984. On Abscesses. Classics in Infectious Diseases. *Rev Infect Dis*, 6(1), pp.122-28.

ONHEALTH. Davis, C.P., 2016. Lung Cancer Symptoms, Stages, and Treatment. [online] Available at: http://www.onhealth.com/content/1/lung_cancer [Accessed 3 Jan. 2016].

PANOS, R.J., MORTENSON, R.L., NICOLLI, S.A., KING JR., T.E., 1990. Clinical deterioration in patients with idiopathic pulmonary fibrosis: Causes and assessment. *The American journal of medicine*, 88(4), pp.396-404.

PATEL, N., JONES, P., ADAMSON, V., THORPE, G., BELCHER, J., SPITERI, M., 2014. P64 Evaluation Of Saliva Biomarkers As Indicators Of health Status And exacerbations in Copd. *BMJ Journals*, 69, pp.A102-A103.

PAZ-ELIZUR, T., KRUPSKY, M., BLUMENSTEIN, S., ELINGER, D., SCHECHTMAN, E. and LIVNEH, Z., 2003. DNA Repair Activity for Oxidative Damage and Risk of Lung Cancer. *CancerSpectrum Knowledge Environment*, 95(17), pp.1312-1319.
PEREIRA, J.M., PAIVA, J.A. AND RELLO, J., 2012. Assessing severity of patients with community-acquired pneumonia. *Seminars in Respiratory and Critical Care medicine*, 33(3), pp.272-283.

PERERA, F.P., 1997. Environment and cancer: who are susceptible?. *Science*, 278, pp.1068-1073.

PEREZ-CORNEJO, P., GOKHALE, A., DURAN, C., CUI, Y., XIAO, Q., HARTZELL, H.C., FAUNDEZ, V., 2012. Anoctamin 1 (Tmem16A) Ca^{2+} -activated chloride channel stoichiometrically interacts with an ezrin-radixin-moesin network. *PNAS*, 109(26), pp.10376-10381.

PHILLIPS, D.H., 1983. Fifty years of benzo(a)pyrene. *Nature*, 303, 468-472.

PIGMAN, W., REID, A.J. 1952. The organic compounds and enzymes of human saliva. *J Am Dent Assoc*, 45, pp.326-38.

RAUS, F.J., TARBET, W.J., MIKLOS, F.L., 1968. Salivary enzymes and calculus formation. *J Periodontal Res*, 3, pp.232-5.

RESPIRATORY CARE and SLEEP MEDICINE, 2016. [online] Available at: http://respiratory-care-sleep-medicine.advanceweb.com/Features/Articles/Early-Screening-for-COPD.aspx [Accessed 1 November 2018].

RESPIRATORY CARE, 2002. AARC Clinical Practice Guideline. *R C Journal*, 37(5), pp.617-625.

RICE, T.W., RUBINSON, L., UYEKI, T.M., VAUGHN, F.L., JOHN, B.B., MILLER, R.R. 3RD, HIGGS, E., RANDOLPH, A.G., SMOOT, B.E., THOMPSON, B.T., NHLBI ARDS NETWORK, 2012. Critical illness from 2009 pandemic influenza A virus and bacterial coinfection in the United States. *Crit Care Med*, 40(5), pp.1487-98.

ROSSI, A., MAIONE, P., COLANTUONI, G., DEL GAIZO, F., GUERREIRO, C., NICOLELLA, D., FERRARA, C. & GRIDELLI, C., 2005. Screening for lung cancer: New horizons?. *Critical reviews in oncology/hematology*, 56, pp.311-320.

Salivary gland, 2016. [online] Available at: https://en.wikipedia.org/wiki/Salivary_gland [Accessed 18 May 2016].

SALTER, S.S., COX, M.J.,TUREK, E.M., CALUS, S.T., COOKSON, W.O., MOFFATT, M.F., TURNER, P., PARKHILL, J., LOMAN, N.J., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, pp.87.

SALVI, S. and BARNES, P.J., 2010. Exposure to Biomass Smoke the Biggest Risk Factor for COPD Globally?. *Chest*, 138(1), pp.3-6.

SCANNAPIECO, F.A., 1994. Saliva-Bacterium Interactions in Oral Microbial Ecology. *Critical Reviews in Oral Biology and Medicine*, 5(3), pp.203-248.

SCHNEYER, L.H., SCHNEYER, C.A., 1964. Secretion of saliva. *Adv Oral Biol*, 1, pp.1-31.

SCHULLER, H.M., MCGAVIN, M.D., ORLOFF, M., RIECHERT, A., PORTER, B., 1995. Simultaneous exposure to nicotine and hyperoxia causes tumours in hamsters. *Lab Invest*, 73(3), pp.448-56.

SEGAL, L.N., ALEKSEYENKO, A.V., CLEMENTE, J.C., KULKARMI, R., WU, B., GAO, Z., CHEN, H., BERGER, K.I., GOLDRING, R.M., ROM, W.N., BLASER, M.J., WEIDEN, M.D. 2013. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome*, 1, pp.19.

SEGATA, N., HAAKE, S.K., MANNON, P., LEMON, K.P., WALDRON, L., GEVERS, D., HUTTENHOWER, C. and IZARD, J., 2012. Composition of the Adult Digestive Tract Bacterial Microbiome Based on Seven Month Surfaces, Tonsils, Throat and Stool Samples. *Genome Biology*, 13(6), pR42.

SHAPIRO, J.A., JACOBS, E.J., THUN, M.J., 2000. Cigar smoking in men and risk of death from tobacco-related cancers. J Natl Cancer Inst, 92, pp.333-337.

SITKI-GREEN, D.L., EDWARDS, R.H., COVINGTON, M.M. AND RAAB-TRAUB, N., 2004. Biology of Epstein-Barr Virus during Infectious Mononucleosis. *J Infect Dis*, 189(3), pp.483-492.

SMITH, G.B., HARPER, P.A., WONG, J.M., LAM, M.S., REID, K.R., PETSIKAS, D., MASSEY, T.E., 2001. Human lung microsomal cytochrome P4501A1 (CYP1A1) activities: impact of smoking status and CYP1A1, aryl hydrocarbon receptor, and glutathione S-transferase M1 genetic polymorphisms. *Cancer Epidemiol Biomarkers Prev*, 10(8), pp.839-53.

SMOKING AND TOBACCO USE, 2016. *CDC - Fact Sheet - Health Effects of Secondhand Smoke - Smoking & Tobacco Use*. [online] Available at: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/secondhand_smoke/health_effects/ [Accessed 29 Dec. 2016].

SOLTANI, M., BAYAT, M., HASHEMI, S.J., ZIA, M., PESTECHIAN, N. 2013. Isolation of *Cryptococcus neoformans* and other opportunistic fungi from pigeon droppings. *J Res Med Sci*, 18(1), pp.56-60.

SOO-QUEE KOH, D. AND CHOON-HUAT KOH, G., 2007. The use of salivary biomarkers in occupational and environmental medicine. *Occup Environ Med*, 64(3), pp.202-210.

SPAGNOLO, P., SCHWARTZ, D.A., 2013. Genetic predisposition to sarcoidosis: another brick in the wall. *European Respiratory Journal*, 41, pp.778-780.

SPIEGELHALDER, B., BARTSCH, H., 1996. Tobacco-specific nitrosamines. *Eur J Cancer Prev*, 1, pp.33-8.

SPRATLIN, J.L, SERKOVA, N.J. and ECKHARDT, S.G. 2009. Clinical Applications of Metabolomics in Oncology: A Review. *Clinical Cancer Research*, 15(2), pp.431-40.

STANOJKOVIC, I., KOTUR-STEVULJEVIC, J., MILENKOVIC, B., SPASIC, S., VUJIC, T., STEFANOVIC, A., ILIC, A., IVANISEVIC, J., 2011. Pulmonary function, oxidative stress and inflammatory markers in severe COPD exacerbation. *Respiratory Medicine*, 105(S1), pp. S31-S37.

STEEL, H.C., COCKERAN, R., ANDERSON, R. and FELDMAN, C., 2013. Overview of Community-Acquired Pneumonia and the Role of Inflammatory Mechanisms in the Immunopathogenesis of Severe Pneumococcal Disease. Mediators of Inflammation, 2013, pp.1-18.

STEEL, H.C., THERON, A.J., COCKERAN, R., ANDERSON, R. AND FELDMAN, C., 2012. Pathogen- and host-directed anti-inflammatory activities of macrolide antibiotics. *Mediators of Inflammation*, 584262, pp.17.

STERLINGCARE, 2017. PARTS OF THE MOUTH. [online] Available at: http://www.sterlingcare.com/resources/diseases-and-conditions-library/view/parts-of-the-mouth/ [Accessed 5 Nov. 2017].

SUGIMOTO, M., SARUTA, J., MATSUKI, C., TO, M., ONUMA, H., KANEKO, M., SOGA, T., TOMITA, M. and TSUKINOKI, K., 2012. Physiological and Environmental Parameters Associated with Mass Spectrometry-Based Salivary Metabolomics Profiles. *Metabolomics*, 9(2), pp.454-463.

SUGIMOTO, M., WONG, D.T., HIRAYAMA, A., SOGA, T., AND TOMITA, M., 2010. Capillary Electrophoresis Mass Spectrometry-Based Saliva Metabolomics Identified Oral, Breast and Pancreatic Cancer-Specific Profiles. *Metabolomics*, 6(1), pp.78-95.

SUZUKI, N., NAKANISHI, K., YONEDA, M., HIROFUJI, T. AND HANIOKA, T., 2016. Relationship Between salivary stress biomarker levels and cigarette smoking in healthy young adults: an exploratory analysis. *Tob Induc Dis*, 14, pp.20.

SZE, M.A., DIMITRIU, P.A., HAYASHI, S., ELLIOTT, W.M., MCDONOUGH, J.E., GOSSELINK, J.V., COOPER, J., SIN, D.D., MOHN, W.W., HOGG, J.C. 2012. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 185, pp.1073-1080.

THE METABOLOMICS LAB, 2014. Metabolomics. [online] Available at: http://www.metabolomics.auckland.ac.nz/index.php/metabolomics [Accessed 25 Jan 2017].

TRACHTENBERG, S., 2005. Mollicutes. *Current Biology*, 15(13), pp.R484. [online] Available at: http://www.cell.com/current-biology/pdf/S0960-9822(05)00684-6.pdf [Accessed 1 Dec. 2017].

TRAVIS, W.D., BRAMBILLA, E., MULLER-HERMELINK, H.K., HARRIS, C.C. and (Eds.), 2004. World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, 2014. Let's Make the Next Generation Tobacco-Free: Your Guide to the 50th Anniversary Surgeon General's Report on Smoking and Health. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. [online] Available at: http://www.surgeongeneral.gov/library/reports/50-years-of-progress/consumer-guide.pdf [Accessed 29 Dec. 2016].

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, 2014. The Health Consequences of Smoking – 50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. [online] Available at: http://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm [Accessed 29 Dec. 2016].

UBHI, B.K., RILEY, J.H., SHAW, P.A., LOMAS, D.A., TAL-SINGER, R., MACNEE, W., GRIFFIN, J.L. AND CONNOR, S.C., 2012. Metabolic profiling Detects Biomarkers of Protein Degradation in COPD Patients. *The European Respiratory Journal*, 40(2), pp.345-55.

US Environmental Protection Agency (EPA). 2009. News Releases by State. Radon Action Could Save Many Lives 01/07/2009. [online] Available at: http://yosemite.epa.gov/opa/admpress.nsf/6424ac1ca a800aab85257359003f5337/8575b629701ce77b8525763800 5ef218!OpenDocument [Accessed 2 Jan. 2017].

VERMEIRE, P.A., RABE, K.F., SORIANO, J.B., MAIER, W.C., 2002. Asthma control and differences in management practices across seven European countries. *Respiratory Medicine*, 96(3), pp.142-149.

VISSE, R., NAGASE, H., 2003. Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases. *Circulation Research*, 92, pp.827-839.

VON HAEHLING, S., LAINSCAK, M., SPRINGER, J., AMKER, S., 2009. Cardiac cachexia: A systematic overview. *Pharmacol Ther*, 121, pp.227-52.

WANG, X., SUN, J., HOLLMANN, R., ZENG, A.P., DECKER, W.D., 2006. Proteomic characterization of density culture of *Bacillus megaterium*. *J Biotechnol*, 126(3), pp.313-24.

WANG, X., Zhang, A., HAN, Y., WANG, P., SUN, H., SONG, G., DONG, T., YUAN, X., ZHANG, M., XIE, N., ZHANG, H., DONG, H., DONG, W., 2012. Urine metabolomics analysis for biomarker

discovery and detection of jaundice syndrome in patients with liver disease. *Mol Cell Proteomics*, 11(8), pp.370-80.

WATANABE, G., 1948. The pathology and physiology of saliva from the standpoint of the acid-base equilibrium. *Igaku*, pp.3-6.

WEBMD, 2016. Pneumonia. [online] Available at: http://www.webmd.boots.com/a-to-z-guides/pneumonia [Accessed 11 Jan. 2017].

WEDZICHA, J.A. AND SEEMUNGAL, T.A.R., 2007. COPD Exacerbations: Defining Their Cause and Prevention. *Lancet*, 370(9589), pp.786-96.

WEI, Q., CHENG, L., AMOS, C.I., WANG, L.-E., GUO, Z., HONG, W.K. and SPITZ, M.R., 2000. Repair of Tobacco Carcinogen-Induced DNA Adducts and Lung Cancer Risk: A Molecular Epidemiologic Study. *Journal of the National Cancer Institute*, 92(21), pp.1764-1772.

WELTE, T., TORRES, A. and NATHWANI, D., 2012. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax*, 67(1), pp.71-79.

WEST, J.B., 1978. Regional differences in the lung. *Chest*, 74(4), pp.426-37.

WICKSTROM, C. and SVENSATER, G., 2008. Salivary Gel-Forming Mucin MUC5B – A Nutrient for Dental Plaque Bacteria. *Oral Microbiology and Immunology*, 23(3), pp.177-82.

WILSON, C.B., JONES, P.W., O'LEARY, C.J., HANSELL, D.M., DOWLING, R.B., COLE, P.J., WILSON, R., 1998. Systemic markers of inflammation in stable bronchiectasis. *Eur Respir J*, 12(4), pp.820-4.

WILSON, M.T., HAMILOS, D.L. 2014. The nasal and sinus microbiome in health and disease. *Curr Allergy Asthma Rep*., 14, pp.485.

WORLD BANK RANKING LIST OF ECONOMIES, 2016. RCGP. Available from: World Bank country distribution renewal July 2016.pdf [12 December 2016].

WORLD CANCER RESEARCH FUND/ AMERICAN INSTITUTE FOR CANCER RESEARCH, 2007. Food, Nutrition, Physical Activity, and the Prevention of cancer: A Global Perspective. [online] Available at: http://www.aicr.org/assets/docs/pdf/reports/Second_Expert_Report.pdf [Accessed 14 March 2016].

WORLD HEALTH ORGANIZATION (WHO) FRAMEWORK CONVENTION ON TOBACCO CONTROL (FCTC), 2017. Decisions at COP7 advance implementation of the WHO Framework Convention on Tobacco Control. [online] Available at: http://www.who.int/fctc/en/ [Accessed 5 Jan. 2017].

WORLD HEALTH ORGANIZATION (WHO), 2010 Burden: mortality, morbidity and risk factors. [online] Available at: http://www.who.int/nmh/publications/ncd_report_chapter1.pdf [Accessed 12 Dec. 2016].

WORLD HEALTH ORGANIZATION (WHO), 2010. Global status report on noncommunicable diseases. [online] Available at: http://www.cancer.org/acs/groups/content/@nho/documents/document/caff2005f4pwsecuredpdf.pdf [Accessed 23 Feb. 2016].

WORLD HEALTH ORGANIZATION (WHO), 2015. Chronic obstructive pulmonary disease (COPD). [online] Available at: http://www.who.int/mediacentre/factsheets/fs315/en/ [Accessed 4 Jan. 2017].

WORLD HEALTH ORGANIZATION INTERNATIONAL AGENCY FOR RESEARCH ON CANCER, 2002. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. [online] Available at: http://monographs.iarc.fr/ENG/Monographs/vol83/mono83.pdf [Accessed 23 Feb. 2016].

WORLD HEALTH ORGANIZATION, n.d. Chronic Respiratory Diseases Asthma: Definition. [online] Available at: http://www.who.int/respiratory/asthma/definition/en/ [Accessed 14 Jan. 2017].

WOTMAN, S., MANDEL, I.D., THOMPSON, R.H., Jr, LARAGH, J.H., 1967. Salivary electrolytes, urea, nitrogen, uric acid, and salt taste threshold in hypertension. *J Oral Ther Pharmacol*., 3, pp.239-50.

WU, Z., DONDEEA, J.B., GHOLAMI, A.M., JANNING, M.C., LEMEER, S., KRAMER, K., ECCLES, S.A., GOLLIN, S.M., GRENMAN, R., WALCH, A., FELLER, S.M., KUSTER, B., 2011. Quantitative Chemical Proteomics Reveals New Potential Drug Targets in Head and Neck Cancer. *Mol Cell Proteomics*, 10(12), pp.M111.011635.

WYNDER, E.L., HOFFMANN, D. 1994. Smoking and lung cancer: scientific challenges and opportunities, *Cancer Res*, 54, pp.5284-5295.

XIA, J., SINELNIKOV, I., HAN, B., AND WISHART, D.S., 2015. MetaboAnalyst 3.0 - making metabolomics more meaningful. *Nucl. Acids Res*., 43, pp.W251-257.

XIAO, H., ZHANG, L., ZHOU, H., LEE, J.M., GARON, E.B. and WONG, D.T.W., 2012. Proteomic Analysis of Human Saliva from Lung Cancer Patients Using Two-Dimensional Difference Gel Electrophoresis and Mass Spectrometry. *Moll Cell Proteomics*, 11(2), pp.M111-012112.

YANG, F., ZENG, X., NING, K., LIU, J. and XU, J., 2012. Saliva Microbiomes Distinguish Caries-Active from Healthy Human Populations. *The ISME Journal*, 6(1), pp.1-10.

ZHANG, A., SHYH-CHANG, N., YANG, H., RAI, A., UMASHANKAR, S., MA, S., SOH, B.S., SUN, L.L., TAI, B.C., NGA, M.E., BHAKOO, K.K., JAYAPAL, S.R., NICHANE, m., YU, Q., AHMED, D.A., TAN, C., SING, W.P., TAM, J., THIRUGANANAM, A., NOGHABI, M.S., PANG, Y.H., ANG, H.S., MITCHELL, W., ROBSON, P., KALDIS, P., SOO, R.A., SWARUP, S., LIM, E.H., LIM, B., 2012**e**. Glycine decarboxylase activity drives non-small cell lung cancer tumor-inititating cells and tumorigenesis. *Cell*, 148(1-2), pp.259-72.

ZHANG, A., SUN, H., HAN, Y., YUAN, Y, WANG, P., SONG, G., YUAN, X., XIE, N., WANG, X., 2012**c**. Exploratory urinary metabolic biomarkers and pathways using UPLC-Q-TOF-HDMS coupled with pattern recognition approach. *Analyst*, 137(18), pp.4200-8.

ZHANG, A., SUN, H., WANG, P., HAN, Y. AND WANG, X., 2012**d**. Modern Analytical Techniques in Metabolomics Analysis. *The Analyst*, 137(2), pp.293-300.

ZHANG, A., SUN, H., WANG, P., HAN, Y., WANG, X., 2012**b**. Recent and potential developments of biofluid analyses in metabolomics. *Journal of Proteomics*, 75(4), pp.1079-1088.

ZHANG, A., SUN, H., WANG, P., WANG, X., 2013. Salivary proteomics in biomedical research. *Clinica Chimica Acta*, 415, pp.261-265.

ZHANG, A., SUN, S., WANG, X., 2012. Saliva Metabolomics Opens Door to Biomarker Discovery, Disease Diagnosis, and Treatment. *Appl Biochem Biotechnol*, 168, pp.1718.

ZHANG, L., XIAO, H., WONG, D.T., 2009. Salivary Biomarkers for Clinical Applications. *Mol Diagn Ther*, 13(4), pp.245-259.

ZHANG, L., XIAO, H., ZHOU, H., SANTIAGO, S., LEE, J.M., GARON, E.B., YANG, J., BRINKMANN, O., YAN, X., AKIN, D., CHIA, D., ELASHOFF, D., PARK, N-H., WONG, D.T.W., 2012**a**. Development of transcriptomic biomarker signature in human saliva to detect lung cancer. *Cellular and Molecular Life Sciences*, 69, pp.3341-3350.

ZUO, F., KAMINSKI, N., EUGUI, E., ALLARD, J., YAKHINI, Z., BEN-DOR, A., LOLLINI, L., MORRIS, D., KIM, Y., DELUSTRO, B., SHEPPARD, D., PADRO, A., SELMAN, M., HELLER, R.A., 2002. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A*, 99(9), pp.6292-7.

7 | APPENDIX

GIG CYMRU NHS WALES | Bwrdd Iechyd Prifysgol Hywel Dda University Health Board

PARTICIPANT INFORMATION SHEET – Version 3 20th February 2017

Title of research programme: "Novel technologies for diagnosing and monitoring pulmonary diseases"

REC reference: 16/WA/0036 **IRAS reference: 187325**

You are being invited to take part in a research study. Before you make a decision it is important to explain why the research is conducted and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Do not hesitate to ask us if there is anything that is not clear or if you would like additional information. Take time to decide whether or not you wish to take part in our study.

1. What is the purpose of the research study?

We believe that many lung diseases can change certain chemicals in people's blood, saliva, urine and chest fluids. These chemicals or 'biomarkers' may be detectable even before a patient develops symptoms or we see changes on hospital scans and chest X-rays. We want to look at new tests including genetic testing in our research laboratories to try and work out which of these chemicals are different in people with lung disease and if these chemicals change with treatment. This research programme is contributing to an educational study, it may help to diagnose lung diseases earlier, and improve our care in the near future.

2. Why have I been chosen to participate?

Your doctor is concerned that you might be suffering from a lung condition. We would like your help by providing occasional extra samples when you attend our hospital for routine tests and treatment.

3. Do I have to participate in this study?

It is up to you to decide whether you would like to take part or not. If you decide to take part, you will be given this Participant Information Sheet and we will ask you to sign a Consent form. You are free to withdraw from the study at any time without providing us with a reason.

4. What will happen to me if I agree to participate in the research?

Firstly we will ask you some questions about your health. We will then ask you to breathe deeply into a tube measuring the amount of carbon monoxide in your breath and spit into a collection container. We may then ask you to provide us with a blood and urine sample.
We may ask to take an extra biopsy or sample of fluid from the lung only if you are undergoing this procedure as part of your routine care. We will take the samples at the same time of this procedure.
It is up to you to decide what sample(s) you want to provide us with, depending on how comfortable you feel.

After the first set of samples, you may be contacted again by letter to ask if you are willing attend a follow up review where the researcher will again ask you questions about your health, a blowing test will be performed and the sputum, blood and urine samples will be collected. This will be timed to coincide with routine appointments wherever possible but may occur at another time more convenient to you.
The samples will be stored in the Hywel Dda University Health Board and sent to different research institutions who are collaborating with the doctors there. These include research laboratories in Aberystwyth University and/or other institutions for specialized laboratory analyses that are not currently available in hospitals. The samples you provide now, will be kept and may also be used for new tests in the future if new technologies are invented.
At present we dispose of all samples after 10 years. However, we are looking at transferring them to a long term local storage tissue bank. We are currently applying for government approval for this longer term storage and if this is granted, your samples will be moved to this facility in the Health Board.
If you want, you can request for your samples to be disposed of at any time.

5. What will I have to do?

There are no lifestyle restrictions if you help. We will want you to continue your normal activities and all usual health treatments and medications will be unchanged.
If you attend hospital at a later date, we may ask you again for samples to see if the 'biomarkers' have changed e.g. after treatment for your lung condition. Again, we would only ask you for biopsy or a lung fluid sample if you were having samples taken as part of your routine hospital treatments.

6. What are the side effects of participating in this study?

There are no known side effects from taking part in this study, but we will need 5 minutes of your time to answer the screening questions and consequently up to 60 minutes do the breath, sputum and blood tests depending on the number of samples you want to provide us with. The extra biopsy or sample of lung fluid will take around 30 extra seconds and will only occur after your routine samples have been taken and if the doctor doing the routine procedure is satisfied that taking an additional sample at the same time is acceptable.

7. What are the benefits associated with me taking part in this research study?

The results of the research will have no direct benefits to you but may benefit patients with lung diseases in the future by moving towards earlier diagnosis and more personalized treatments. Unfortunately, we cannot pay you for participating in this study.

8. What will happen if something goes wrong?

If you feel that you have any reason to complain about any aspect of the way you have been approached in the hospital or further treated during the course of the study, the normal National Health Service complaints mechanisms are available to you.

9. Will my participation in this study be kept confidential?

All the collected information about you during the course of our research programme will be kept strictly confidential. Any information that leaves the hospital will be coded so you cannot be identified from it. In addition, we will not give any identifiable information to life insurance, private medical insurance companies or any other third parties.

10. What will happen with the results generated by this research programme?

The findings from our study may be published in scientific journals and presented at conferences in a fully anonymized way. You are welcome to contact the researchers for a report every 2-3 years.

11. Who is organizing and funding this research?

This research is the result of a collaboration between Hywel Dda University Health Board and Aberystwyth University. It is funded by the European Social Fund, the Brazilian Government through the project "Science without Borders," Tenovus Charity and a pharmaceutical company ProTEM Services Ltd. The company has had no influence on the design of this research and will not keep samples but it jointly funds a KESS PhD student within the University.

12. Who has reviewed the study?

Our research study has undergone review by medical doctors and scientific researchers within Hywel Dda University Health Board and across Wales As well as an NHS Research Ethics Committee (WALES REC 7).

Contacts for further information:

For independent advice, please contact Mr Chris Tattersall, R&D Department, Withybush Hospital, Tel: 01437 773813.
Email: chris.tattersall@wales.nhs.uk

If you have any further queries please do not hesitate to contact:

Prof. Keir Lewis
Chief Investigator
Hon. Consultant, Hywel Dda University Health Board &
Professor of Respiratory Medicine, Swansea University
Telephone: 01554 783133
Email: k.e.lewis@swansea.ac.uk

Supplementary Figure 3.1 | Participant Information Sheet utilised for patient recruitment at Bronglais, Glangwili and Prince Philip General Hospitals.

Patients attending weekly respiratory clinics were screened by lung consultants. The Participant Information Sheet was then posted to potential eligible patients prior to their visits to the hospital. This process was done in order to allow patients enough time to reflect on their decision regarding their participation in the study and consult their family members.

Supplementary Table 3.1 | COPD Group

Patients affected by COPD classified by gender and age group in accordance with the smoking status and smoking pack years mean as well as the FEV1% predicted.

COPD Group					
		Gender		FEV1% Predicted Mean	
		Male	Female	Male	Female
Age group	20s	1	-	-	-
	30s	-	-	-	-
	40s	1	-	51	-
	50s	4	2	50.6	43
	60s	10	6	38.6	62.5
	70s	12	7	74.5	61
	80s	3	1	44	43
Smoking Status	**Current Smoker**	8	4		
	Smoking Pack Years Mean	31.9	51.5		
	Ex-Smoker	20	9		
	Smoking Pack Years Mean	34.4	45.2		
	Never	2	3		

Supplementary Table 3.2 | Other Diagnoses Group

Patients suffering of various respiratory diseases including idiopathic pulmonary fibrosis, bronchiectasis, asthma, pneumonia and sarcoidosis were analysed as a whole group named 'Other' based on their gender, age group and smoking status.

Other Diagnoses Group			
		Gender	
		Male	Female
Age group	20s	-	-
	30s	-	1
	40s	2	-
	50s	2	2
	60s	5	5
	70s	2	-
	80s	3	3
Smoking Status	**Current Smoker**	1	1
	Smoking Pack Years Mean	-	-
	Ex-Smoker	7	4
	Smoking Pack Years Mean	39.2	25
	Never	6	6

Supplementary Table 3.3 | Lung Cancer Group

Patients with confirmed lung cancer diagnoses classified by gender and age group according to their smoking status and smoking pack years mean.

Lung Cancer Group			
		Gender	
		Male	Female
Age Group	20s	-	-
	30s	-	-
	40s	-	-
	50s	-	1
	60s	4	1
	70s	1	-
	80s	1	-
Smoking History	**Current Smoker**	1	1
	Smoking Pack Years Mean	53	40
	Ex-Smoker	4	1
	Smoking Pack Years Mean	50	75
	Never	-	-

Supplementary Table 3.4 | Healthy Control Group

Table composed of healthy individuals that were currently smoking or former smokers at the time of the recruitment. The table has been constructed based on the individuals' gender, age group and smoking status.

Healthy Control Group			
		Gender	
		Male	Female
Age group	20s	1	3
	30s	2	1
	40s	3	5
	50s	7	3
	60s	5	2
	70s	1	1
	80s	3	-
Smoking Status	**Current Smoker**	6	2
	Smoking Pack Years Mean	54	50
	Ex-Smoker	10	1
	Smoking Pack Years Mean	25.3	10
	Never	6	-

Supplementary Table 3.5 | Summarised Frequent Respiratory Comorbidities, Drug and Medical Histories Found in Recruited Patients from the COPD, LC and Other Groups
Summarised patient information detailing frequent clinical data such as comorbidities, drug history, use and medical history. Steroids and inhalers were most common prescribed, along with antibiotics.

Condition	Drug History	Use	Medical History
Respiratory	Carbocisteine	Mucolytic	COPD, Asthma, Pneumonia, Bronchiectasis, Interstitial lung disease
	Fostair	Inhaler	
	Mucodyne	Carbocisteine to clear sputum in the respiratory tract	
	Salbutamol/ Ventolin	Opens up medium and large airways in the lungs	
	Seretide	Management of Asthma and COPD	
	Prednisolone	Steroid	
	Spiriva/ Tiotropium	Bronchodilator	
	Symbicort	Inhaler	

Supplementary Table 3.6 | Summarised Frequent Cardiac Comorbidities, Drug History, Use and Medical History in Recruited Patients from the COPD, LC and Other Groups
Summarised clinical data information from patients suffering of COPD, LC or other respiratory diseases with the most frequent cardiac comorbidities. Attached there is also the commonest prescribed cardiac medication, its use and common medical history.

Condition	Drug History	Use	Medical History
Cardiac	Bisoprolol	Heart Diseases, Chest pain	Coronary Artery Bypass Graft, Heart Failure, Hypertenstion
	Clopidogrel	Heart Diseases, Stroke	
	Furosemide	Fluid build-up due to Heart Failure, Liver Scarring	
	Losartan	High blood pressure	

Supplementary Table 3.7 | Summarised General Comorbidities in Recruited Patients Suffering of Respiratory Diseases

Summarised clinical data information of comorbidities, from patients recruited for the COPD, LC and Other groups with several conditions, drug history, use and medical history.

Condition	Drug History	Use	Medical History
Orthopedic	Calcium with Vit. D	Bone supplements	**Arthritis, Osteoporosis**
Neurological	Gabapentin	Epilepsy, Partial seizures	**Epilepsy, Neuropathic Pain, Hot flashes**
Abdominal	Lansoprazole	Inhibits stomach's production of acid	**Irritable Bowel Syndrome**
	Omeprazole	Gastroesophageal reflux disease	**Hepatitis, Diabetes Type II**

Supplementary Table 4.1 | Nextera XT Index Primer1 (N7##) and Nextera XT Index Primer2 (S5##) from Illumina

Nextera index primers used in the 16S rRNA Amplicon Sequencing analysis for the sequence construction (Table adapted after Cameron, 2015).

Sample ID		Nextera XT Barcode			
		Index 1 (i7)	Sequence	Index 2 (i5)	Sequence
A001	COPD	N701	TAAGGCGA	S517	GCGTAAGA
A004	HC	N702	CGTACTAG	S517	GCGTAAGA
A030	COPD	N703	AGGCAGAA	S517	GCGTAAGA
A035	COPD	N704	TCCTGAGC	S517	GCGTAAGA
A038	Other	N705	GGACTCCT	S517	GCGTAAGA
A039	LC	N706	TAGGCATG	S517	GCGTAAGA
A041	Other	N707	CTCTCTAC	S517	GCGTAAGA
A042	COPD	N708	CAGAGAGG	S517	GCGTAAGA
A046	HC	N709	GCTACGCT	S517	GCGTAAGA
A048	HC	N710	CGAGGCTG	S517	GCGTAAGA
A052	HC	N711	AAGAGGCA	S517	GCGTAAGA
A054	COPD	N712	GTAGAGGA	S517	GCGTAAGA
A055	COPD	N701	TAAGGCGA	S502	CTCTCTAT
A058	COPD	N702	CGTACTAG	S502	CTCTCTAT
A059	HC	N703	AGGCAGAA	S502	CTCTCTAT
A061	COPD	N704	TCCTGAGC	S502	CTCTCTAT
A062	COPD	N705	GGACTCCT	S502	CTCTCTAT
A063	COPD	N706	TAGGCATG	S502	CTCTCTAT
A067	COPD	N707	CTCTCTAC	S502	CTCTCTAT
A071	COPD	N708	CAGAGAGG	S502	CTCTCTAT
A075	COPD	N709	GCTACGCT	S502	CTCTCTAT
A077	COPD	N710	CGAGGCTG	S502	CTCTCTAT

A078	HC	N711	AAGAGGCA	S502	CTCTCTAT
A079	HC	N712	GTAGAGGA	S502	CTCTCTAT
A086	HC	N701	TAAGGCGA	S503	TATCCTCT
A087	HC	N702	CGTACTAG	S503	TATCCTCT
A088	HC	N703	AGGCAGAA	S503	TATCCTCT
A089	COPD	N704	TCCTGAGC	S503	TATCCTCT
A090	LC	N705	GGACTCCT	S503	TATCCTCT
A091	HC	N706	TAGGCATG	S503	TATCCTCT
A093	HC	N707	CTCTCTAC	S503	TATCCTCT
A094	LC	N708	CAGAGAGG	S503	TATCCTCT
A095	HC	N709	GCTACGCT	S503	TATCCTCT
A101	COPD	N710	CGAGGCTG	S503	TATCCTCT
A105	COPD	N711	AAGAGGCA	S503	TATCCTCT
A109	COPD	N701	GTAGAGGA	S504	AGAGTAGA
A110	COPD	N702	TAAGGCGA	S504	AGAGTAGA
A112	COPD	N703	CGTACTAG	S504	AGAGTAGA
A114	HC	N704	AGGCAGAA	S504	AGAGTAGA
A115	COPD	N705	TCCTGAGC	S504	AGAGTAGA
A116	COPD	N706	GGACTCCT	S504	AGAGTAGA
A117	HC	N707	TAGGCATG	S504	AGAGTAGA
A118	COPD	N708	CTCTCTAC	S504	AGAGTAGA
A120	HC	N709	CAGAGAGG	S504	AGAGTAGA
A121	Other	N710	GCTACGCT	S504	AGAGTAGA
A122	HC	N711	CGAGGCTG	S504	AGAGTAGA
A123	HC	N712	AAGAGGCA	S504	AGAGTAGA
A124	HC	N701	GTAGAGGA	S505	GTAAGGAG
A125	HC	N702	TAAGGCGA	S505	GTAAGGAG
A126	COPD	N703	CGTACTAG	S505	GTAAGGAG
A127	HC	N704	AGGCAGAA	S505	GTAAGGAG
A129	COPD	N705	TCCTGAGC	S505	GTAAGGAG
A131	COPD	N706	GGACTCCT	S505	GTAAGGAG
A133	COPD	N707	TAGGCATG	S505	GTAAGGAG
A134	LC	N708	CTCTCTAC	S505	GTAAGGAG
A137	COPD	N709	CAGAGAGG	S505	GTAAGGAG
A138	HC	N710	GCTACGCT	S505	GTAAGGAG
A141	HC	N711	CGAGGCTG	S505	GTAAGGAG
A145	COPD	N712	AAGAGGCA	S505	GTAAGGAG
A147	COPD	N701	GTAGAGGA	S506	ACTGCATA
A149	HC	N702	TAAGGCGA	S506	ACTGCATA
A161	COPD	N703	CGTACTAG	S506	ACTGCATA
B002	COPD	N704	AGGCAGAA	S506	ACTGCATA
B005	COPD	N705	TCCTGAGC	S506	ACTGCATA
B011	Other	N706	GGACTCCT	S506	ACTGCATA
B012	COPD	N707	TAGGCATG	S506	ACTGCATA
B014	COPD	N708	CTCTCTAC	S506	ACTGCATA
B015	HC	N709	CAGAGAGG	S506	ACTGCATA
C004	COPD	N710	GCTACGCT	S506	ACTGCATA

C009	COPD	N711	CGAGGCTG	S506	ACTGCATA
C010	COPD	N701	TAAGGCGA	S517	GCGTAAGA
C011/C004	COPD	N702	CGTACTAG	S517	GCGTAAGA