THE BACTERIAL LUNG TISSUE MICROBIOME IN THE PATHOGENESIS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by

MARC ALEXANDER SZE

B.M.L.Sc., The University of British Columbia, 2009;

M.Sc, The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2015

Abstract

Rationale: Several laboratories have shown that the decline in lung function in Chronic Obstructive Pulmonary Disease (COPD) is associated with increased formation of tertiary lymphoid follicles. This provides direct histological evidence in support of the hypothesis that the decline in lung function is associated with activation of an adaptive immune response. The antigens responsible for driving this immune activation remain poorly understood. The recent realization that the human lung contains a bacterial microbiome that changes in association with the presence of COPD suggests the hypothesis that bacteria arising from within this microbiome might be responsible for activating the adaptive immune response in COPD.

Approach: The research described in this thesis examines the lung tissue bacterial microbiome from patients with mild to moderate COPD as well as patients with very severe COPD. The bacterial microbiome from these studies utilized either nested or touchdown PCR followed by 454^{TM} pyrotag sequencing of specific variable regions on the 16S rRNA gene. Changes in the microbiome were examined in relation to histological estimates of emphysematous destruction of the lung and inflammatory immune cell infiltration associated with this tissue remodeling process. Finally, *Haemophilus influenzae,* a bacterium identified from this microbiome, known to cause inflammation was compared to the host tissue repair process.

Results: The different bacterial community was present in control and mild (GOLD 1) compared to moderate (GOLD 2) COPD. The community composition was also different between donor lung tissue and very severe (GOLD 4) COPD. Further, the analysis identified a list of 10 OTUs that discriminated between lung tissue affected by GOLD 4 COPD and controls. In addition, the

data presented here indicate that the host immune response to these organisms precedes the structural changes associated with COPD.

Conclusion: Collectively, these data confirm that there is a small but diverse microbiome in the normal human lung that becomes less diverse in COPD. Furthermore, the disappearance/appearance of certain OTUs can discriminate between control and COPD affected lung tissue and that some of these OTUs are associated with the inflammatory immune cell infiltration and tissue destruction that occurs in COPD.

Preface

This research was approved by the UBC-Providence Health Care Research Ethics Board. The certificate number for these projects fall under H10-00843.

The introduction section from "what is the microbiota" onwards was published in the International Journal of Chronic Obstructive Pulmonary Disease.
Marc Sze, James Hogg, and Don Sin. (2014) Bacterial microbiome of lungs in COPD. Int J Chron Obstruct Pulmon Dis. 9:229-38. doi: 10.2147/COPD.S38932. eCollection 2014.
I conducted all the research and wrote the entire first draft of the manuscript. Subsequent drafts were edited by James Hogg and Don Sin.

The chapter on the host response to the bacterial microbiome COPD was previously published in the American Journal of Respiratory and Critical Care Medicine.
Marc Sze, Pedro Dimitriu, Masa Suzuki, John McDonough, Joshua Campbell, John Brothers II, John Erb-Downward, Gary Huffnagle, Shizu Hayahsi, Mark Elliott, Joel Cooper, Don Sin, Marc Lenburg, Avrum Spira, William Mohn, and James Hogg. The host response to the lung microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2015,epub ahead of print DOI: 10.1164/rccm.201502-0223OC.
I performed all the microbiome experiments and data analysis as well as wrote the first draft. Subsequent drafts were edited by James Hogg and Don Sin and submitted after approval from all authors.

The chapter on 16S bacterial quantification with ddPCR was previously published in PLOS One. Marc Sze, Meysam Abbasi, James Hogg, and Don Sin. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD. PLOS One. 2014, 9(10): e110351. doi:10.1371/journal.pone.0110351.

I conducted all the research and wrote the entire first draft of the manuscript with Meysam Abbasi. Subsequent drafts were edited by James Hogg and Don Sin.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

(Alphabetical Order)

16S: 16 Svedberg Units

µg: microgram

µL: microliter

µM: micromolar

µm: micrometer

AC: Adenocarcinoma

AE: Elution Buffer

AIDS: Acquired Immune Deficiency Syndrome

Alv: Alveolar Tissue

ANOVA: Analysis of Variance

BAC: Bronchioalveolar Carcinoma

BAL: Bronchoalveolar Lavage

Bp: Base Pair

BSC 2: Biosafety Cabinet Level 2

C.difficile: Clostridium difficile

CD: Cluster of Differentiation

CF: Cystic Fibrosis

CLE: Centrilobular Emphysema

cm: centimeter

COPD: Chronic Obstructive Pulmonary Disease

CS: Carcinoma

CT: Computed Tomography

CV: Coefficent of Variation

DAVID: Database for Annotation, Visualization and Integrated Discovery

ddPCR: Droplet Digital Polymerase Chain Reaction

DLCO: Lung Diffusing Capacity

DNA: Deoxyribonucleic Acid

DNase: Deoxyribonuclease

dNTP: Deoxyribonucleotide triphophate

F: Female

FDR: False Discovery Rate

FEV_1: Functional Expiratory Volume in 1 second

FVC: Functional Vital Capacity

fSAD: functional Small Airways Disease

EB: EDTA-Tris buffer

E.coli: Escherichia coli

E.meningoseptica: Elizabethkingia meningoseptica

GEO: Gene Expression Omnibus

H&E: Hematoxylin and Eosin

HIV: Human Immunodeficiency Virus

HLI: Heart Lung Innovation Centre

HMW: High Molecular Weight

HPD: Protein D

HRCT: High Resolution Computed Tomography

GOLD: Global initiative for Chronic Obstructive Lung Disease

ID: Identity

IL: Interleukin

LC: Lung Components

LCC: Large Cell Carcinoma

Lm : Mean Linear Intercept

MDCT: Multi Detector Computed Tomography

MicroCT: Micro Computed Tomography

MMP: Matrix Metalloproteinase

mMRC: Modified Medical Research Council Dyspnea Scale

M: Male

n: number

N/A: Not Applicable

NHLBI: National Heart, Lung and Blood Institute

NSCC: Non-Small Cell Carcinoma

OCT: Optimal Cooling Temperature

OOB: Out-of-bag

OTU: Operational Taxonomic Unit

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

PCoA: Principle Coordinates Analysis

PERMANOVA: Permutational Multivariate Analysis of Variance

PLE: Panlobular Emphysema

PMN: Polymorphonuclear Cell

qPCR: Quantitative Polymerase Chain Reaction

RMA: Robust Multichip Average

RNase: Ribonuclease

ROS: Reactive Oxygen Species

Rpp40: Ribonuclease P Protein Subunit p40

S: OTU Richness

SA: Surface Area

S.pneumonia(e): Streptococcus pneumonia(e)

Spp: species

RCCA: Regularized Canonical Correlation Analysis

RF: Random Forests

rRNA: Ribosomal Ribonucleic Acid

P: P-value

TCRA: T-Cell Receptor Alpha

TNF: Tumor Necrosis Factor

V: Variable Region

Vv: Volume Fraction

WHO: World Health Organization

UBC-PHC: University of British Columbia – Providence Health Care

Acknowledgements

First, I would like to thank both my supervisors, Drs James Hogg and Don Sin, for all their support throughout my PhD. They have shown enormous patience and trust as the project unfolded and new directions were explored. I could not have asked for better supervisors and their advice has helped keep me focused. Without them I would not have been able to complete this thesis in three years.

Second, I would like to extend my gratitude to the members of my research committee. Both Drs Bill Mohn and Del Dorscheid have had useful comments and advice as to how this thesis should be structured and written. Their support throughout this process and their willingness to help me reach my goals was something I greatly appreciated. I want to also specifically thank Dr. Bill Mohn, without his guidance I would have been lost at sea, so to speak, analyzing and trying to interpret the data we generated on the bacterial microbiome.

Third, I would like to thank Dr. Pedro Dimitriu, without his guidance and support I would never have gained the confidence needed to be able to analyze the bacterial microbiome. His patience is explaining and breaking down how the tools work and are used was invaluable and essential in the completion of this thesis.

Fourth, I would like to thank Dr John McDonough for all his help with the microCT processing and analysis as well as Dr. Masa Suzuki for his help in the quantitative histology processing and analysis. Without these two individuals this project would not have been completed as quickly as it was.

Fifth, I would like to thank all the members from both the Hogg and Sin laboratory. They have been very helpful and supportive throughout the years that I spent completing my thesis. Because of them the wet laboratory portion of the experiments that I did went smoothly without any major problems.

This thesis was the result of a large number of collaborations and I would like to finally thank all those from other institutions who have helped make this thesis possible. Without you the information presented here would not have been produced.

Dedication

To Tammy, Mom, Dad, and family

Chapter 1: Background

1.1 Chronic Obstructive Pulmonary Disease (COPD)

1.1.1 Burden of COPD in the World:

A recent report has now placed COPD in the top 10 for years of life lost and top 3 for causes of death globally (1). This recent study used real world data and emphasizes the point that previous projections made back in 2006 and 2008 may have been too conservative in their estimate of COPD being the 4th leading cause of death by 2030 (2,3). Furthermore as smoking rates remain constant and still very high in certain parts of the world (4–6) the contribution of COPD to all cause mortality is projected to continue to rise while those of diseases like HIV/AIDS and respiratory infections are projected to continue to fall (2,3). The increased prevalence of COPD represents a very real global health problem which ranges from increased comorbidities, including heart disease (the leading cause of death worldwide), asthma, lung cancer, depression, and child health impairment (in the form of increased susceptibility to respiratory infections) (7,8), to increased cost of hospitalizations (9,10); with the cost for treatment in North America being more than some of the figures quoted internationally (11). Recent findings that even in a normal population COPD can be detected at a reasonable rate, 6.6% for GOLD 1 (mild COPD) and 5.6 % for GOLD 2 and above (moderate to severe) (12), further emphasizes the importance of identifying viable treatment options and increased research into the pathogenesis of the disease itself. Clearly, with younger and larger smoking populations in other parts of the world (6,13), along with under diagnosis and mismanagement in some of these countries (14), this disease will only see increased importance on the world stage for years to come.

1.1.2 COPD Diagnosis

The primary risk factor for developing COPD is cigarette smoking (15), but other factors including the burning of biomass fuel for cooking and heat (16,17), atmospheric pollution from the exhaust of internal combustion engines and industrial processes (18), and the genetic background of the subjects (19), of which alpha-1 anti-trypsin is the most widely reported (20,21), are also involved. The diagnosis of COPD is primarily based on criteria set out by the Global Initiative for Chronic Obstructive Lung Disease (GOLD), initially sponsored by the World Health Organization (WHO) and the Heart, Lung and Blood Institute (NHLBI) in the United States. These guidelines are updated on an annual basis and currently recommend that a diagnosis of COPD requires a reduction in the post bronchodilator FEV_1/FVC ratio to below 0.7 (22–24). The severity can be classified into four grades (from 1-4) based on measurements of FEV_1 expressed as a percentage of its predicted value for each subject [Table 1] (24). The most recent modifications of the GOLD guidelines now include measurements of quality of life, exacerbation history, and symptoms (24). This has redefined the categories according to GOLD group A – D (24) in which group A and C have fewer symptoms and group B and D have more symptoms [Figure 1]. Although this newer classification system is designed to improve management, and provide better estimates of the effect of treating symptoms (25), it does not appear to add value to simple spirometry in predicting mortality (26,27). Further investigation of the relationship between symptoms and mortality are clearly needed (28). In addition there is urgent need for simple tests capable of predicting what Fletcher and colleagues termed "the susceptible minority of smokers". This minority develop the rapid rate of decline in FEV_1 that leads to severe (GOLD 3) and very severe (GOLD 4) grades of COPD. Some of these tests may

include improvements in the quantitative analysis of inspiratory and expiratory High-Resolution Computed Tomography (HRCT) scans (29–32) as well as blood tests for predictive biomarkers of the decline in FEV_1 (33).

Table 1: Breakdown of Cutoffs for GOLD Grades of Disease

GOLD Grade	FEV_1	FEV_1/FVC
GOLD 1	$\geq$ 80% predicted	< 0.7
GOLD 2	50% $\leq$ FEV_1 < 80% predicted	< 0.7
GOLD 3	30% $\leq$ FEV_1 < 50% predicted	< 0.7
GOLD 4	FEV_1 < 30% predicted	< 0.7

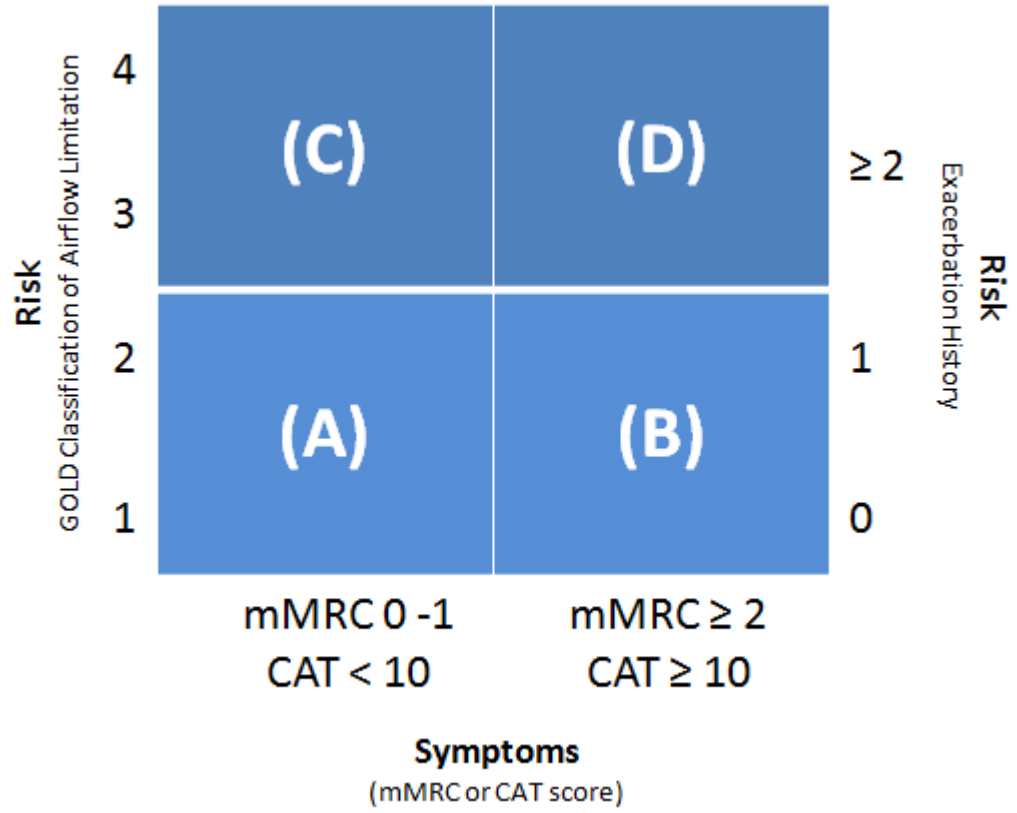


Figure 1: Overview of GOLD A, B, C, and D Categories

1.1.3 The Different Phenotypes of COPD

COPD is an umbrella term that encompasses individuals who suffer from variable combinations of chronic bronchitis, emphysema, and/or small airways disease (34).

Chronic bronchitis is defined by the presence of a chronic productive cough that occurs daily for at least three months of the year for two successive years (35,36). It is also associated with mucus hypersecretion, epithelial remodeling, and alteration of airway surface tension (34,37).

There is also histological evidence of chronic inflammation and enlargement of bronchial mucus glands as well as goblet cell metaplasia of the epithelial lining of the cartilagenous airways (i.e. , bronchi) in the lower respiratory tract. The goblet cell metaplasia normally starts at the trachea and main stem bronchi and continues all the way down to airways that measure approximately 3 mm in diameter (38). Additionally the increased mucus production, estimated from cough and sputum production, correlates with an increased inflammatory immune cell infiltration and remodeling of muscle, connective tissue, and microvasculature in these airways (39,40).

Emphysema is defined by enlargement of the airspaces with destruction of the alveolar lung tissue and can be divided into either a Centrilobular (CLE), Panlobular (PLE), or Paraseptal phenotype (41,42). Before the introduction of computed tomography (CT), diagnosis of emphysema and the classification of these 3 basic phenotypes of destruction was based on the examination of post mortem lungs in a fixed inflated state (43,44). Clinical diagnosis of emphysema, before the availability of CT scans, was based on symptoms of dyspnea, wheezing, and signs of hyperinflation of the lung caused by advanced destruction of lung tissue and the formation of emphysematous bullae (45). The presence of less severe emphysema in inflated lung specimens, resected as treatment for lung cancer, demonstrated that emphysema was sometimes present in smokers without COPD and this observation has been repeatedly confirmed in living patients since the introduction of CT scans (46,47). Furthermore, several studies now indicate that this early appearance of emphysema in patients with normal lung function may predict those most likely to show the rapid decline in FEV_1 that leads to severe COPD. The centrilobular phenotype (CLE) is the most common form of emphysema observed in smokers and is characterized by destruction of respiratory bronchioles in the centriacinar

regions of the lung lobule with the initial preservation of the more distal alveolar ducts and sacs (42). However, the natural progression of CLE leads to the destruction of the entire lung lobule with eventual coalescence of many destroyed lobules to form bullous lesions. The emphysema seen in CLE is typically heterogenous and tends to involve the upper lobes to a greater degree than the lower lobes (48,49). The panlobular phenotype (PLE) of emphysematous destruction produces much more uniform destruction of the entire lung lobule and is commonly but not exclusively observed in alpha-1 antitrypsin deficiency (42). Although PLE is normally distributed across the whole lung it is predominately found in the lower lobes (49). Paraseptal emphysema represents the third separate phenotype that destroys the outer portions of the acini leaving the center intact (50). It has been implicated in the pathogenesis of pneumomothorax in young adults (51) and is often found in association with centrilobular emphysematous destruction in the middle aged and elderly.

The final pathologic phenotype of COPD is the airway dominant disease in which there is little if any emphysematous destruction. The rate at which the gas exchanging tissues of the lung fill and empty, is determined by the resistance to flow offered by the airways that conduct air to the gas exchanging tissue and the elastic properties of these tissue that stores energy during their expansion, that is used to drive air out of the lung. Moreover as the product of the units of resistance to flow through the airways (cm H2O/litre/second) and the compliance (cm H2O/litre) of the alveolar tissue simplifies to time; increases in either the resistance to flow produced by obstruction in the small airways or compliance produced by emphysematous destruction of lung elastic recoil lengthen the time required to empty the lung. Furthermore, when the time required to empty the lungs exceeds the maximum time between breaths, gas will remain trapped within

the lungs and the vital capacity will be reduced (52). Although the small conducting airways (less than 2mm in internal diameter) offer very little resistance to airflow in normal adult lungs, they become the major site of obstruction to airflow in COPD (53). The introduction of microCT made it possible to identify and count the total number of terminal bronchioles / lung. Using this approach provided evidence that COPD is associated with a massive destruction of terminal bronchioles, before emphysema can be detected by microCT (54) examination, and long before the emphysematous lesions become large enough to be visualized by HRCT (55–58). The qualitative analysis of MDCT scans taken at full inspiration and expiration has made it possible to estimate where the gas was trapped at the end of a forced expiration. Moreover the application of parametric response mapping allows for the discovery of regions of the lung that have functional small airways disease (fSAD). This was done by registering the voxels of scans taken on full inspiration to those present on full expiration in order to discover the regions of lung that trapped gas excessively on full expiration, but were emphysema free on inspiration. Additionally, the ability to repeat this analysis over time in the same person has made it possible to show regions of the lung that initially had functional small airways disease develop emphysema over time (59).

Work from several laboratories has shown that the pathology found in the small conducting airways increases in association with a decline in FEV_1. These changes include an increased infiltration of the airway wall tissue and lumen with innate inflammatory cells, that include polymorphonuclear leukocytes (PMN) and marcophages, as well as cells that participate in the adaptive immune response, like CD4+ T-cells and B cells (53). This increase in inflammatory immune cells has also been attributed to an increase in inflammatory cytokines (IL6, IL1β, TNF-

α, interferon-γ, etc.) (60). These cytokines can recruit other cells into the lungs like CD8+ T-cells, which have been found to be increased in COPD (61). These CD8+ T-cells have the capacity to release elastolytic enzymes such as MMP2, MMP9, and MMP12 capable of degrading lung tissue (62). Additionally, as CD4+ T-cells and B-cells increase with COPD severity so do the number of tertiary lymphoid follicles associated with the small airways (53). This observation suggests that there may be an adaptive immune response to an antigen that potentially drives the progression of disease (63). Although the innate immune response is important in COPD it is possible that the adaptive immune response, predominantly driven by CD4+ T-cells and B-cells, may orchestrate the persistent inflammation seen in COPD and the destruction mediated by both CD8+ T-cells and neutrophils.

Acute exacerbations of COPD are defined by an increase in symptoms over baseline everyday secretion, production, purulence, viscosity, or volume of sputum. This can be accompanied with nasal discharge (runny nose), sore throat, fever, and increased coughing or wheezing (64). Expert opinion suggests that these symptoms generally need to last for at least 2 days in order to constitute an exacerbation (65). The new guidelines recently created for diagnosis of an acute exacerbation of COPD states that there should be continuous worsening of an individual's condition from their usual stable state (24,66). However, there are still multiple definitions or interpretations of what an acute exacerbation in COPD is and recent studies have found that only around 20% of clinical trials use symptom based definitions for acute exacerbations (67). Bacteria and viruses have long been associated with exacerbations and have been found to be the major cause of most exacerbations in COPD (68–70). Infectious agents (whether they are virus,

or bacteria) account for close to 80% of all causes of acute exacerbations. Additionally, air pollution can worsen these symptoms and increase the need for hospitalization (71,72).

1.1.4 Potential Inflammatory Initiator of COPD

A decline in FEV_1 can be associated with increased inflammatory and immune cell infiltration of the lung tissue. This infiltration has been shown to be a mixed population of inflammatory immune cells that include macrophages, CD4+ and CD8+ T-cells, and B cells (73). Very recently a study comparing lung tissue from smokers with the "usual" centrilobular phenotype of emphysematous destruction to lung tissue from patients with the panlobular emphysematous phenotype of COPD had the same infiltrating immune and inflammatory profiles (74). This new data suggests the hypothesis that the elastase anti-elastase imbalance created by A1AT deficiency could cause disease by producing abnormal elastin fragments that act as autoantigens capable of driving an adaptive immune response in A1AT deficiency (75,76). Alternatively, the reduction in diversity observed within the microbiome of patients with COPD might either create conditions that allow new species of organisms to emerge and produce infection or present new microbial antigens that stimulate the host immune system to respond.

The recognition that the decline in lung function associated with COPD is correlated with increased exacerbation frequency is well established in the literature (77), as is the concept that exacerbations are primarily caused by either viruses or bacteria (78). Moreover a recent report from CanCOLD shows that the symptoms of exacerbations are frequently observed in subjects without COPD (79). This finding supports an old clinical adage that the common cold, that is an

annoyance to persons with healthy lungs, may be a life threatening event in a person whose lung defenses are severely compromised by COPD. These observations are relevant to the vicious cycle hypothesis where a background of impaired lung defense causes increased microbial colonization, which leads to an acute inflammatory response, leading to increased destruction of the lung which never fully recovers before the next infection/colonization by a bacteria or virus (78). Moreover they also fit with certain aspects of the Dutch hypothesis where a hyperreactive host response is thought to contribute to the pathogenesis of COPD and asthma as well as the British hypothesis that implicated mucus hypersecretion to the symptoms of both asthma and COPD.

Although there is substantial evidence for important overlaps between asthma and COPD (80), there are still attributes that are quite different between the two (81). One of the key differences is that asthma is dominated by a T-helper 2 lymphocyte (Th2) cell response as well as increases in eosinophil and mast cell populations within the lungs (82). COPD cell infiltration into the lung, in contrast, is typified by increases in neutrophils, macrophages, and cytotoxic T-cells (82). CD4+ T-cells and B-cells have also been found to be substantially increased in COPD (53). This milieu is more usual of the T-helper 1 lymphocyte (Th1) response. Although there are differences, synergistic worsening of symptoms does occur when both diseases are present. First, in a 20 year study it was shown that asthmatics had the highest hazard and attributable risk factor for developing COPD (80). Second, declines in lung function have been shown to be more severe when the smoker is asthmatic (83). Third, a family history of asthma and smoking can increase the probability that an individual will have COPD (84).

So how does COPD relate to the bacterial microbiome within our lungs? Can changes to this identified microbiome have any impact on the progression of disease or are these communities merely bystanders in the overall pathogenesis? Before I attempt to show data to help provide some answers to these questions it is first important to review what the microbiota is and what has already been published both on the bacterial microbiome in lungs and in COPD.

1.2 The Bacterial Microbiome[1]

1.2.1 What is the Microbiota?

In general, the microbiota consists of microorganisms that inhabit a particular site or place (e.g., the gastrointestinal tract, skin, lung, etc.) (85–87). These microorganisms can consist of, but are not limited to, bacteria, viruses, and fungi (88,89), of which the most widely studied is the bacteria. In the human body bacteria out number our human cells by a factor of 10 (90). Most of these bacteria are found in the gastrointestinal (GI) tract (90). Other sources of habitation include the mouth, nose, and skin (91). Bacteria co-exist with fungi and viruses in these locations; however, their role and function within these eco-systems are now only recently being established (92–94). The first large scale human microbiome studies were conducted on the gut (95,96). The initial hypothesis was that most microorganisms residing in the GI tract could pose health threats for humans (97). However, with careful investigation, it became clear that many bacteria in the GI tract were beneficial in providing protection against death and disease (98). For example, some bacteria are required for the production of essential micro-nutrients (such as

[1] This section has been previously published in the International Journal of Chronic Obstructive Pulmonary Disease. Sze MA, Hogg JC, Sin DD. Bacterial microbiome of lungs in COPD. *Int J Chron Obstruct Pulmon Dis*. 2014 Feb; 9:229-38. Doi: 10.2147/COPD.S38932.eCollection.

vitamin K), protection against pathogens (such as *C.difficile*), and regulation of the host inflammatory responses (99,100), and disease phenotypes (101–103). It is now recognized that perturbations in the gut microbiome may be responsible for a wide range of diseases including pseudomembranous colitis (104), inflammatory bowel disease (105), and even non-GI conditions such as obesity and cardiovascular diseases (106).

1.2.2 The BAL, Bronchial Brushing, and Endotracheal Lung Microbiome

Studies of the lung microbiome are now just emerging. Hilty et al. were the first to show that the lungs were not sterile and bacteria are found in the lower airways (107). Using clone libraries, they interrogated the 16S rRNA gene fragments for bacterial communities in healthy control subjects (n=8), patients with asthma (n=11), and those with moderate to severe COPD (n=5). They demonstrated that bronchoalveolar lavage (BAL) fluid and bronchial brushings contained different bacterial communities to those found in the nasal cavity or the oropharynx (107). More specifically, they showed that there was increased representation of *Proteobacteria* in the COPD and asthmatic airways, which was accompanied by a reduction in *Bacteriodetes* in the COPD samples (107). This was the first study to suggest that there was a unique bacterial community in the lungs, which may change with disease.

Erb-Downward et al. evaluated the bacterial microbiome in non-smokers and smokers with normal lung function (101). This study evaluated mostly BAL fluid, complemented by lung tissue samples, which were obtained from patients with very severe COPD (101). They found that there was no significant difference in the overall bacterial community composition between

non-smokers, healthy smokers, and COPD patients (101). However, they showed that there was significant heterogeneity and diversity in the bacterial micorbiome across different regions of the same lung (101).

Huang et al., extended these findings by using a bacterial 16S phylo chip to determine bacterial composition of endotracheal aspirates, which were obtained in a small number of intubated patients with severe COPD (108). Interestingly, however, they noted two distinct and divergent bacterial populations in the COPD samples (108). One group of COPD patients demonstrated a loss of diversity in their bacterial composition, similar to what was reported by Hilty et al. previously (108). The second group, on the other hand, showed increased diversity of community composition and in particular an in the increase number of bacteria in the *Firmicute* phylum (108). They hypothesized that disease progression in COPD was associated with greater bacterial diversity and increased airway representation of *Firmicutes* (108).

These observations were supported by data generated with resected lung tissue specimens of patients with very severe COPD, which demonstrated increased representation of *Firmicutes* (109). As with other studies (101,103,109), this study also showed that there was no difference in the bacterial community composition in the lungs between smokers and non-smokers. Finally, although the total bacterial load was much lower in the lung tissue samples compared with BAL fluid, neither the tissue samples nor BAL fluid have shown any significant differences in the total bacterial load between COPD patients and healthy (control) subjects (107,109).

Pragman et al. extended these prior studies by determining the lung microbiome in a small group of control subjects and in patients with moderate to severe COPD using BAL fluid samples. They found that the bacterial communities of COPD lungs were distinct from those of normal lungs, though there were no significant differences across disease severity (102). This was the first study that used BAL fluid to evaluate the lung microbiome across GOLD (Global initiative for Chronic Obstructive Lung Disease) grades of severity. These data suggest that those with COPD, regardless of severity, have a different bacterial microbiome in their lungs compared with those who do not have COPD and that changes in the microbial communities occur very early in the disease process. Importantly, they also showed that there was segregation of bacterial communities according to the use (or non-use) of inhaled corticosteroids or bronchodilators (102). However, since this study was cross-sectional, causality could not be ascribed (102).

Together, the studies to date suggest that the changes in the lung microbiome in COPD occur early in the disease process and remain relatively stable with disease progression. Finally, although there is a lack of uniformity on the organisms found in the lung microbiome of COPD and normal lungs, most studies have reported both an increased representation of bacteria in the *Firmicute* phylum as well as an overall numeric dominance of the *Proteobacteria* phylum in the COPD lungs.

One limitation of these studies is that they focused solely on the DNA component. Therefore, it is possible that many of the bacteria studied are in fact dead. What needs to be done in the COPD field are studies that have been done by researchers in the cystic fibrosis field and

compare the communities identified by the DNA approach to those that can be found by an RNA approach (110,111). This importance in sorting out the living bacteria from the dead bacteria is even more important due to studies that show a large proportion of sequences that are identified do not belong to living bacteria (112). This comparison would help in sorting out what may be dead cells from living cells. Further, there are other approaches that can be used on DNA samples to limit the contribution that dead bacteria will have on the sample (113,114).

Traditionally it is believed that the airways below the vocal cords are sterile (101–103,107–109,115). It has been well documented that the mucociliary transport system is a key component in keeping the level of bacteria in the lungs low (116). Numerous reports have shown that a dysfunctional mucociliary transport can lead to infection (117–119). Further, macrophages and neutrophils clear any microorganism or particle that evades this transport system (120,121). Although many studies have found bacterial sequences within the lower airways caution needs to be used when discussing whether these bacteria truly live in this environment. Based on the historical literature there is wide acceptance that bacteria get into the lower airways on a regular basis (122) and are cleared by normal processes (116,120). These DNA based approaches are a good first step but to truly prove that the lung is not "sterile" will require more than just sequencing studies since these studies may just be detecting bacteria that normally "invade" airways but are later cleared without causing any harm to the host. As well, changes in the bacteria community that gets into the lung could still be just as important to the progression of COPD.

1.2.3 The Lung Tissue Bacterial Microbiome

The first analysis of the bacterial lung tissue microbiome in COPD was performed by Sze et al., which largely confirmed the previous findings of Hilty et al. Both groups showed that there was a significant difference in the bacterial community composition detected between COPD and "normal" (control) lungs (109). Notably, they found that the total number of bacteria in lung tissue, which has a bacterial density of 10-100 bacterial cells per 1000 human cells, was relatively small as compared with the gut microflora (109). However, even at this low concentration, Sze et al found using two separate techniques (terminal restriction length polymorphism analysis and pyrotag sequencing) that the lungs of patients with very severe COPD contained a different community of bacteria than those of controls or patients with cystic fibrosis (109). Using indicator species analysis, they noted that these differences were largely driven by bacteria belonging to either the *Proteobacteria* or *Firmicute* phylum (109), which was consistent with what was previously reported by Hilty et al. and Huang et al. However, by using lung tissue samples rather than bronchoscopic specimens (which are prone to upper airway contamination), Sze et al provided the first evidence that lungs of COPD patients harbored a distinct microbiome (more on this in the subsequent section).

In a separate study done in cystic fibrosis, Goddard et al showed that the microbial diversity within the upper airways was significantly greater than what could be found in the lower airways of explanted tissue (123). This means that many studies based on sputum may actually over – estimate the bacterial diversity as well as bacterial density that actually gets into the lower airways and lung tissue.

1.2.4 The Initial Location of the Bacterial Lung Microbiome

To date most studies of lung bacterial microbiome data have been generated using BAL samples. However, because the bronchoscope has to traverse through the upper airways, these data may be confounded by contamination of organisms in the mouth or the nose (115,124). To surmount this limitation Charlson et al instituted several quality measures in their bronchoscopic techniques including rinsing of mouth with an antispectic solution prior to bronchoscopy, restraints on suctioning through the bronchoscope in the upper airways during the procedure, and discarding of initital BAL samples (124). Similar to previous studies (which did not implement these stringent quality measures for bronchoscopy), they found that the overall bacterial load was higher in the BAL samples when compared with the negative controls (124) and that there was good concordance of the bacterial community across the samples (124).

From other lung disease research the major source of bacteria for the lung microbiome is thought to be from the upper airways (125,126). In COPD this notion is supported in part by a recent study conducted by Segal et al (115). They found that some healthy individuals carried organisms in the BAL fluid that were commonly observed in the supraglottic region, whereas other individuals demonstrated unique organisms in the BAL fluid that were not found in the upper airways (115). Interestingly, individuals in whom there was substantial overlap in the microbiome between BAL fluid and the upper airway demonstrated increased lung inflammation, characterized by increased lymphocytes and neutrophils in the BAL fluid, compared to those whose BAL microbiome was distinct from that of the upper airways. These results suggest that “contamination’ of bacteria from the supraglottic region into the lungs may

elicit an inflammatory response in the lungs. These data raise the possibility that bacteria from the mouth may alter the normal lung microbiome, contributing to "disease". However, it is also possible that these data were confounded by contamination of microbial flora during bronchoscopy.

1.2.5 Can the Oral Bacterial Microbiome Play a Role in COPD?

Few studies have evaluated possible changes in the bacterial communities of the oral cavity of smokers as compared to non-smokers (103,127). Charlson et al. found that there was indeed a difference in the oral microbiome between smokers and non-smokers, most notably in the *Firmicute* phylum (127). This was supported by Morris et al. who also showed using a much larger sample size that there were detectable differences in the oral microbiome between smokers and non-smokers (103). Both studies found differences in the representation of *Neisseria* species (103,127). They also showed that many bacteria in the oral cavity can be found in the lungs. However, some bacteria (such as *Enterobacteriaceae*, *Haemophilus*, *Methylobacterium*, and *Ralstonia*), which are found in both areas, are enriched in the lungs as compared with the oral cavity (103). The substantial overlap in the microbiome between the oral cavity and lungs may be related to micro-aspiration (122). It is well known that nearly all individuals micro-aspirate during sleep (122). However, aspirated bacteria are cleared by an intact mucociliary clearance system during the day, which prevents pneumonia (128). In individuals with COPD there is an impairment of this mucociliary clearance system (129). This impairment could lead to mucus hypersecretion, pooling of mucus and mucus plugging in the airways (130), entrapping these aspirated bacteria in the lungs and causing them to acclimate and grow in this new ecosystem.

This process may also elicit a local immune response, contributing to the persistent lung inflammation observed in COPD airways (even following smoking cessation). It is possible that these "aspirated" bacteria may stimulate the formation of the tertiary lymphoid follicles, which are prominent in the small airways of patients with very severe COPD. The impaired mucociliary clearance and cilia (131) in COPD lungs may also permit the entry and growth of non-commensal bacterial pathogens in the lung causing acute worsening of symptoms and exacerbations. The persistence of non-commensal organisms, coupled with the ongoing inflammatory response, may shift the lung microbiome in COPD (108,109).

1.2.6 The Microbiome and Inflammation in COPD

Very little research has been done to investigate the role the microbiome plays in inflammation with respect to COPD. One of the first studies to look at the potential role the microbiome may have in inflammation in COPD was by Segal et al (115) and was previously discussed. A more detailed study on specific bacteria within the lung microbiome was done by Larsen et al (132). In this study they tested the bacteria identified previously (107) to be important in asthma or COPD against dendritic cells. What they found was that commensal bacteria had lower cytokine expression of IL-23, IL-12p70, and IL-10 than bacteria believed to be pathogenic (*Haemophilus spp* and *Moraxella spp.*) (132). They also went on to show that *Prevotella spp,* a bacterium identified as a commensal, could reduce the amount of *Haemophilus influenza*-induced IL-12p70 (132). Others have studied *Lactobacillus* and their role in reducing airway inflammation in mouse models of asthma (133,134). At this point in time these types of studies have found that in general commensal organisms can have an overall anti-inflammatory effect while non-

commensals can have a pro-inflammatory effect. Caution needs to be taken with these results since what is commensal in one body site may be pathogenic in another (135). These data show promise in identifying both bacteria that could be important in dampening the immune response and those that could accentuate it. Ultimately, both *in vivo* and *in vitro* studies investigating how different bacteria can drive the tertiary lymphoid follicle formation and increased infiltration of B-cells and CD4 T-cells into the lungs of those with COPD are what need to be accomplished to push this area forward.

1.2.7 What We Have Learned So Far on the Lung Microbiome in COPD

Overall, our understanding of the lung bacterial microbiome in COPD is still in its infancy. Despite excitement about the lung microbiome, there remain inconsistencies in data and poor reproducibility of findings across studies. For example, although the first few studies to investigate COPD (mostly using BAL or sputum samples) have found that the bacterial diversity decreases as disease worsens (101,107), subsequent studies using lung tissue samples have failed to show significant differences in bacterial diversity (109). One possible explanation for this conflict is that tissue samples contain mostly parenchyma (mixed with airways and blood vessels), while BAL and sputum samples mostly reflect the airways [Table 2], which could result in a greater airway to alveolar sampling admixture. This could suggest that the bacterial microbiome within the airways is different from those in the alveolar tissue giving rise to different micro-niches within different compartments of the lung. It should also be noted that the concept of reduced diversity with disease progression has not been consistently replicated even

in studies using BAL samples (102). Contrary to earlier studies, one recent study showed that diversity paradoxically increased in severe COPD as compared with controls (102).

One common finding so far from the COPD studies is the increased abundance of bacteria in the *Firmicute* phylum in moderate, severe, and even very severe disease (102,107,109), with a few notable exceptions (101,136). Erb-Downward et al found in patients with very severe disease that the predominant bacteria in lung tissue were those from the *Proteobacteria* phylum (101). This finding is similar to what has been found by Hilty et al. Huang et al. provide a plausible explanation to reconcile these differences. They speculated that there are two types of bacterial microbiomes related to COPD, one that is dominated by *Proteobacteria* and the other that is dominated by *Firmicutes* and which is associated with increased diversity (108). Further studies will be needed to investigate and resolve this controversy and determine the role of the lung microbiome in disease progression of COPD.

Although we can draw knowledge from the cystic fibrosis and bronchiectasis literature, the role the bacterial microbiome plays in those two disease process is more than likely to be very different. There are some similarities between the diseases with respect to the bacterial microbiome. For example, in COPD and other lung diseases a loss of diversity can be associated with worse disease. However, in COPD this does not necessarily lead to an outgrowth of single organisms as it does in cystic fibrosis and a corresponding loss in community evenness. This difference along with the low bacterial density in COPD versus cystic fibrosis implies that bacteria potentially have very different roles in either disease progression. Many of these studies are based in DNA sequencing approaches and may only identify bacteria that simply pass

through the lungs yet the type of bacteria that the lung is exposed to, even at low bacterial density, may have a large impact on inflammation and progression in COPD.

Table 2: Breakdown of the different bacterial microbiome studies in COPD

Study	Controls	COPD	Sampling Method	Predominant Stage of COPD
Hilty et al.[115]	8	5	Bronchial Brush	2-3
Erb-Downward et al.[109]	10	4	BAL*	1
	0	6	Tissue	4
Huang et al.[116]	0	8	Endotracheal Aspirates	Exacerbation
Sze et al. [117]	16	8	Tissue	4
Pragman et al. [110]	10	22	BAL	2

*Abbreviation: BAL = bronchoalveolar lavage

These previous studies have been observational survey studies of the bacterial microbiome between control and COPD. Those that focused on more mild disease did not find any difference between COPD and control (102). The studies that focused on more severe disease found measurable shifts in the bacterial microbiome between control and COPD (107,109). These data suggest that the bacterial microbiome may not change until later GOLD grade, specifically at moderate COPD. Additionally, these survey studies have not tried to tackle the question of whether or not this bacterial microbiome or specific bacteria within it could be an active facilitator in the progression of disease. This thesis aims to first pinpoint at what GOLD grade bacterial community composition changes can be detected in lung tissue and how changes in this bacterial microbiome could potentially influence and drive the disease pathogenesis of COPD.

Chapter 2: Experimental Approach

2.1 Working Hypothesis

There is a detectable effect of the bacterial microbiome or specific bacteria within this microbiome on chronic obstructive pulmonary disease (COPD) pathogenesis.

2.2 Specific Aims

1) Determine if changes in the bacterial microbiome found in very severe (GOLD 4) COPD can also be seen in mild and moderate COPD versus control lung tissue.

2) Determine if the bacterial microbiome as well as specific bacteria in this microbiome correlate with structural changes that occur in COPD.

3) Determine if the bacterial microbiome as well as specific bacteria in this microbiome correlate with inflammatory and immune cell changes in COPD.

4) Determine the role of a specific bacterium from aims 2 and 3 and whether it has different effects on adaptive immune activation and immune cells.

Chapter 3: The Bacterial Microbiome in Mild and Moderate COPD

3.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is currently the 4th leading cause of death worldwide (2). A loss of terminal bronchioles as well as a robust inflammatory process that involves the innate and adaptive immune system has been characterized in disease (53,57,61,137). Additionally, increased infiltration of macrophages, CD4+ T-cells, and B cells has been correlated with emphysematous tissue destruction, as measured by the mean linear intercept (Lm) (138,139). The correlation between Lm and these infiltrating inflammatory cells occurs before emphysematous destruction can be detected by regular MDCT scans (57,140). However, the target of this adaptive immune response is not known. Some possible reasons involve either an autoimmune response to structural components like elastin (141) or environmental responses to viruses or bacteria (142–144).

Recent work on the bacterial microbiome has focused largely on COPD GOLD 4 disease (101,108,109). A few studies have examined the bacterial microbiome in more mild disease (102,107,115), however their results are mixed. Although there are some marked differences between control samples and COPD (102,107) the differences between various GOLD grades are not clear. Further, there is some controversy over whether contamination from either the mouth or nose is a key contributor to the sequences analyzed for the bacterial microbiome. Studies have shown that both BAL and bronchial brush samples can have many potential sources of contamination, including bacteria from the mouth (115,124). Thus many of the bacteria

sequenced may not identify the bacteria that are reflective of the communities found in the lower respiratory tract and alveolar tissue of the lungs. However, newer studies that use similar protocols show that contamination due to the mouth bacterial microbiome is minimal (136,145,146). Even if this latest result proves to be true, many of the currently published studies analyze a relatively small sample size. Additionally, there has not yet been a study on the lung tissue bacterial microbiome in mild and moderate COPD.

This current study has the largest number of independent tissue samples studied for the bacterial microbiome in mild and moderate COPD. The main hypothesis tested was that the bacterial microbiome is different among control, mild, and moderate COPD. It also investigated whether early inflammation, in particular macrophages, B cells, and CD4+ T-cells, could be detected before noticeable increases in Lm occur and whether these infiltrating inflammatory cells are potentially driven by either bacterial communities or specific bacteria within the microbiome.

3.2 Methods

3.2.1 Tissue Preparation and Extraction

Lung tissue was obtained from the tissue registry at St. Paul's Hospital. Ethics approval was obtained for this study from the University of British Columbia - Providence Health Care (UBC-PHC) research ethics board. Informed consent was obtained, through a written consent form, and approved by the UBC-PHC research ethics board for patients who underwent lung resection therapy for various pulmonary conditions for collection and use in this study. Lung tissue from the tumor-free part of the resected lung segment was used. Three individuals (one from the

control group, two from the COPD group) had used inhaled corticosteroids, and none had symptoms of an acute respiratory tract infection documented in the two weeks prior to surgery. Individuals from whom the lung tissue cores were obtained had the following diagnoses: adenocarcinoma (n = 13), squamous cell carcinoma (n = 12), large cell carcinoma (n = 9) and other (n = 6) [Table 3]. The other category is made up of both non-malignant and rarer types of cancer.

Table 3: The Breakdown of the Different Tumor Types from Resection Therapy

	Adenocarcinoma	Squamous Cell	Other
Control	11 (39%)	3 (11%)	14 (50%)
GOLD 1	4 (19%)	10 (48%)	7 (33%)
GOLD 2	5 (20%)	11 (44%)	9 (36%)

Resected lung tissues were inflated with cryomatrix (OCT) at constant pressure (30cm H_2O) and then frozen in liquid nitrogen. 2 cm thick contiguous transverse slices were then made and tissue samples were taken from one of these slices. From the same core, consecutive frozen sections were cut on a cryomicrotome and were assigned as follows: sections 1–5, 8–12, 14–18 were allocated for qPCR or microbiome analysis, and sections 6–7, 13, 19–20 were allocated for histological and immunohistochemical staining. This sectioning protocol was repeated in quintuplicate for each lung tissue core and at least two tissue cores were examined from each patient specimen. Samples for histology were stained for CD4+ T-cells, CD8+ T-cells, B-cells, Neutrophils (PMN), Macrophages, and Eosinophils. A hematoxylin and eosin (H&E) and Movat pentachrome stain were used to quantify elastin, airway wall thickness, and mean linear intercept (Lm). Table 4 lists the breakdown of staining for the different cell types. Optimal staining concentrations were determined from a serial dilution of each antibody and both a positive

control (tissue known to contain the cell of interest) and a negative control (staining protocol without the target antibody) were performed.

Table 4: Immunohistochemical Staining of Airway Inflammatory Cells

Antibody name	Antibody Type	Host Species	Against	Company	Catalog #	Clone Name	Dilution	Pre-treatment
CD68+ (Macrophage)	Monoclonal	Mouse	Human	DAKO	M0718	EBM11	1/75	acetone 10 min. at room temperature
NK1+ (Natural Killer Cell)	Monoclonal	Mouse	Human	DAKO	M1014	DAKO-NK1	1/50	acetone 10 min. at room temperature
CD79α+ (B-lymphocyte)	Monoclonal	Mouse	Human	DAKO	M7050	JCB117	1/50	acetone 10 min. at room temperature
CD4+ (Helper-inducer T-lymphocyte)	Monoclonal	Mouse	Human	DAKO	M0716	MT310	1/100	acetone 10 min. at room temperature
CD8+ (Cytotoxic T-lymphocyte)	Monoclonal	Mouse	Human	DAKO	M7103	C8/144b	1/100	acetone 10 min. at room temperature
Neutrophil Elastase (Neutrophil)	Monoclonal	Mouse	Human	DAKO	M752	NP57	1/100	acetone 10 min. at room temperature

* Hansel stain used for Eosinophils

3.2.2 Study Design

In total there were 28 individuals in the control group (normal lung function as measured by spirometry), 21 individuals in the GOLD 1 group and 25 individuals in the GOLD 2 group. DNA was extracted for both qPCR and 454^{TM} pyrotag sequencing [Figure 2]. The qPCR was performed to determine the total bacterial load at two lung heights within the same individual. The 454^{TM} pyrotag sequencing was used to compare the bacterial community composition (measured by both the alpha diversity (the number of species and evenness (147)) and beta diversity (the turnover of species (147)). The bacterial community composition was then compared to the data on the inflammation and emphysematous tissue destruction (Lm) to identify potential individual bacteria or community measures that were associated with either a protective or destructive role in COPD. Both PCR controls (negatives) and DNA extraction controls (extraction negatives) were sequenced with the above protocol. The extraction negatives were water controls subjected to the same DNA extraction process and nested PCR as the tissue samples. The negatives were subjected to the same nested PCR as the tissue samples.

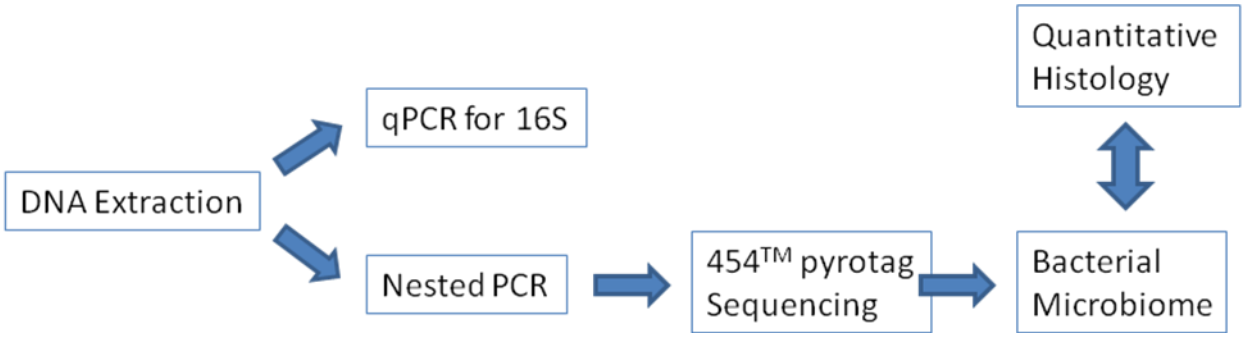


Figure 2: Workflow of the Different Components of the Study on Mild and Moderate COPD.

3.2.3 Quantitative Histology

Digital images of these slides were obtained using the Aperio ePathology slide capture system (Aperio systems, Newark, NJ, USA) and analyzed using Image-Pro Plus software (Media Cybernetics, Rockville, MD, USA) to obtain volume fractions (Vvs) occupied by each of the infiltrating cells.

The volume fraction was obtained by using a 40x magnification and subsequent counting of the ratio of positive cells to total tissue. A grid was placed on the entire section and each box on the grid was assigned a number. A random number generator was used to select 4 locations where a 40x magnification image was to be taken and quantified. Total positive cell counts needed to be approximately 200 cells and extra images were obtained if the 4 images were not sufficient. A grid based system was used for counts and only tissue or positive cells that fell on these points were counted. For alveolar tissue measurements the airways and blood vessels were excluded from the random number generator selection process.

Lm was measured using a standardized grid with lines (1 mm in length) that surrounded the entire image. Airways and blood vessels were excluded so any lines crossing these structures were not counted. After this exclusion the number of times alveolar tissue crossed the lines (intercepts) was counted. Lm was then calculated using the following equation:

Lm = [(Number of lines) x (line length)] / (Number of intercepts)

Airway wall thickness was measured from the volume of the lumen being subtracted from the total volume of the airway and dividing by the area of the basement membrane. This is depicted in equation form below.

Airway wall thickness = [(Volume of Airway) – (Volume of Lumen)] / (Area of the Basement Membrane)

3.2.4 QPCR

The 16S rRNA gene assay standard curve was based on a serial dilution of genomic DNA extracted from *Escherichia coli* JM109. The *E.coli* JM109 was grown on LB medium agar plates and DNA was extracted on a pooled sample of 7 medium sized colonies. In order to obtain a cell count of the sample the average size of the *E.coli* JM109 genome of 4.5 million base pairs was converted to Daltons (660 Daltons per base pair). The Dalton measurement was then converted to grams / cell using the DNA concentration (ng/µL) of the extracted genomic DNA of *E.coli* JM109 to obtain a cell / µL value. A single dissociation curve was observed for both the 16S rRNA gene assay at 84-85°C. The cycling conditions for the 16S rRNA gene were previously published [117]. Modifications to this existing protocol included normalization to µg of DNA instead of to the human housekeeping gene Rpp40. This was done since both gave similar results with respect to finding no difference between the 16S bacterial load between controls and disease. Further running a single plate rather than two plates decreased the time that needed to be spent on performing 16S quantification.

The standard curves for the 16S rRNA gene assays were $y = -3.2543x + 29.03$, $R^2 = 0.99$, with an efficiency of 102.9% and $y = -3.2896x + 29.526$, $R^2 = 0.99$, with an efficiency of 101.4%. The average of all the negative control samples were subtracted from the sample average and this value was then multiplied by 7 to generate a 16S copy number and then divided by 5 to generate a 16S/µL value. This was then divided by the ng of DNA/µL and multiplied by 1000 to generate the 16S copies / µg of DNA.

3.2.5 454™ Pyrotag Sequencing

HotstarTaq DNA Polymerase, 10x PCR Buffer, and dNTP mix from Qiagen (Maryland, USA), along with primers from SIGMA (Missouri, USA), and RNase and DNase free water was used for all reactions. The exact volume in a single tube was 5µL of 10x PCR Buffer, 1µL dNTP mix 2µL of the forward primer, 2µL of the reverse primer, 0.25 µL of HotstarTaq DNA polymerase, 34.75µL of RNase and DNase free water, and 5µL of the template DNA. All the PCR reactions were carried out on a Bio-Rad MyCycler Thermal Cycler (Ontario, Canada).

A nested PCR approach was utilized to generate a 550bp product, which spanned the V1-V3 region of the 16S rRNA gene [117], used for the sequencing. The first round of PCR consisted of the following cycling conditions:

[95°C for 15 minutes] x 1
[94°C for 40 seconds, 57°C for 30 seconds, 72°C for 1 minute and 30 seconds] x 40

Using the forward primer 27F (5'- AGAGTTTGATCMTGGCTCAG) and reverse primer 907R (5'- CCGTCAATTCMTTTGAGTTT) to generate an 881bp product. A second round of PCR

consisted of a forward primer (5'-AGAGTTTGATCMTGGCTCAG) and a reverse primer (5'-GWATTACCGCGGCKGCTG) at the following cycling conditions:

[95°C for 15 minutes] x 1
[94°C for 40 seconds, 61°C for 40 seconds, 72° for 1 minute] x 40
[72°C for 10 minutes] x 1

Barcodes were included with each primer to allow for the samples to be sequenced in a pooled library. This consisted of two half runs utilizing 84 unique primer and barcode combinations within each run [Table 5]. The fusion sequence needed to attach the amplicons to the 454™ beads was CCATCTCATCCCTGCGTGTCTCCGACTCAG for the forward primers and CCTATCCCCTGTGTGCCTTGGCAGTCTCAG for the reverse primer.

Table 5: List of Barcodes Used for 454™ Pyrotag Sequencing

Name	Unique Barcode	Name	Unique Barcode	Name	Unique Barcode
MID1	ACGAGTGCGT	MID30	AGACTATACT	MID57	CGCGTATACA
MID2	ACGCTCGACA	MID31	AGCGTCGTCT	MID58	CGTACAGTCA
MID3	AGACGCACTC	MID32	AGTACGCTAT	MID59	CGTACTCAGA
MID4	AGCACTGTAG	MID33	ATAGAGTACT	MID60	CTACGCTCTA
MID5	ATCAGACACG	MID34	CACGCTACGT	MID61	CTATAGCGTA
MID6	ATATCGCGAG	MID35	CAGTAGACGT	MID62	TACGTCATCA
MID7	CGTGTCTCTA	MID36	CGACGTGACT	MID63	TAGTCGCATA
MID8	CTCGCGTGTC	MID37	TACACACACT	MID64	TATATATACA
MID10	TCTCTATGCG	MID38	TACACGTGAT	MID65	TATGCTAGTA
MID11	TGATACGTCT	MID39	TACAGATCGT	MID66	TCACGCGAGA
MID13	CATAGTAGTG	MID40	TACGCTGTCT	MID67	TCGATAGTGA
MID14	CGAGAGATAC	MID41	TAGTGTAGAT	MID68	TCGCTGCGTA
MID15	ATACGACGTA	MID42	TCGATCACGT	MID69	TCTGACGTCA
MID16	TCACGTACTA	MID43	TCGCACTAGT	MID70	TGAGTCAGTA
MID17	CGTCTAGTAC	MID44	TCTAGCGACT	MID71	TGTAGTGTGA
MID18	TCTACGTAGC	MID45	TCTATACTAT	MID72	TGTCACACGA
MID19	TGTACTACTC	MID46	TGACGTATGT	MID73	TGTCGTCGCA
MID20	ACGACTACAG	MID47	TGTGAGTAGT	MID74	ACACATACGC
MID21	CGTAGACTAG	MID48	ACAGTATATA	MID75	ACAGTCGTGC
MID22	TACGAGTATG	MID49	ACGCGATCGA	MID76	ACATGACGAC
MID23	TACTCTCGTG	MID50	ACTAGCAGTA	MID77	ACGACAGCTC
MID24	TAGAGACGAG	MID51	AGCTCACGTA	MID78	ACGTCTCATC
MID25	TCGTCGCTCG	MID52	AGTATACATA	MID79	ACTCATCTAC
MID26	ACATACGCGT	MID53	AGTCGAGAGA	MID80	ACTCGCGCAC
MID27	ACGCGAGTAT	MID54	AGTGCTACGA	MID81	AGAGCGTCAC
MID28	ACTACTATGT	MID55	CGATCGTATA	MID82	AGCGACTAGC
MID29	ACTGTACAGT	MID56	CGCAGTACGA	MID83	AGTAGTGATC

3.2.6 Pipeline to Generate Samples

Samples were run on a 1% agarose gel to confirm the presence of the 881bp product. Samples that contained the product continued onto the second PCR while those that did not contain the product had the first PCR round redone until an 881bp product was detected on the 1% agarose

gel. Upon confirmation of a product in the first PCR, 5µL of this PCR was used as a template for the second PCR. These samples were then purified using the Beckman-Coulter Agencourt Ampure 5mL Kit system (Ontario, Canada) and eluted out in Qiagen EB buffer. Samples were assessed for quality and quantity using a Nanodrop. A 1% agarose gel was run to confirm a single product at approximately 550bp and that no smaller products remained after the purification. The samples were stored in a -20ºC freezer and thawed once to transfer in strip tubes to Genome Quebec.

3.2.7 Data Analysis

The Vv of each cell and tissue type present within the bronchiolar and alveolar tissue of the stained histological sections were inserted into a multi-level cascade sampling design to compute the accumulated volume of infiltrating inflammatory immune cells. To assess correlations between quantitative histological measures linear mixed-effects models ('lme' function of the R *nlme* package) were used and correlations with an FDR < 0.1 were considered significant. Linear mixed-effect models were chosen to account for the fact that multiple samples were obtained from the same individual and not truly independent of one another.

The total number of reads for each community (tissue samples, negative controls, and extraction negative controls) was normalized, using random sub-sampling, to 3302 the smallest number of reads among the samples after denoising. The OTU abundance table was also filtered to exclude OTUs with a cumulative summed abundance of ≤5 reads. These steps were done to control for differences in sequencing depth before alpha diversity and community similarity analyses. The

extraction negative controls were analyzed in the same way as the samples and all downstream tests were designed to explore differences between both the GOLD grade groups and controls and differences in the extraction negative controls and the GOLD and control samples. Community similarity was visualized with non-metric multidimensional scaling analysis (NMDS) of pair-wise Bray-Curtis dissimilarities computed from square-root transformed OTU relative abundances. The effects of disease status, were statistically examined with a permutational multivariate analysis of variance (PERMANOVA; (148), which enabled the quantification of the relative proportion of variability explained by each source of variation in the model (149). Ordinations and PERMANOVA were performed using the vegan package in R and R studio software (150).

To identify OTUs that discriminate between control and GOLD 4 communities, we used the Random Forests (RF) algorithm, an ensemble-based supervised classification method that generates multiple weak classifier decision trees (151). The classification error rate was measured by out-of-bag (OOB) estimation for each group. An importance measure was calculated for each feature (OTU) based on the loss of accuracy in classification when the OTU was removed from analysis. The importance measure was then determined using the Boruta package, a feature selection algorithm built around the RF algorithm (152). RF and Boruta analyses were performed in R and R studio.

In order to compare between the bacterial microbiome and quantitative histological measurements within the lung tissue regularized canonical correlation analysis (RCCA)(19,153) was utilized. Networks were generated using the built in functionality of the mixOmics R

package (153). For all linear mixed-effects analysis, values with an FDR < 0.1 were considered significant. The RCCA does not generate significance values commonly used (e.g. P-value, or T statistic).

3.3 Results

There was no significant difference between the controls, GOLD 1, and GOLD 2 groups with respect to age and sex. There was also no difference in the smoking histories of the three groups ($P > 0.05$). There was a significant difference in the measured lung function of all three groups with controls having the highest lung function and GOLD 2 having the lowest [Table 6] ($P < 0.05$). Data on antibiotic usage on these specific samples was not well documented.

Table 6: Clinical Characteristics of the Sample Groups (average ± SD)

	Controls (n=28)	**GOLD 1** (n=21)	**GOLD 2** (n=25)
Age	65.7 ± 9.6	66.0 ± 8.9	63 ± 9.2
Sex (M:F:Unknown)	16:11:1	14:7:0	17:8:0
Smoking History (pack-years)	44.8 ± 31.1	48.0 ± 25.2	47.3 ± 27.8
Smoking Status (Never:Current:Ex:Unknown)	1:14:10:3	0:12:7:2	0:17:5:6
FEV_1/FVC	77.4 ± 4.9	64.3 ± 4.3*	62.0 ± 7.0**
FEV_1 (percent predicted)	100.0 ± 12.5	89.9 ± 9.0†	69.0 ± 6.6**

* P<0.0001 between controls versus GOLD 1
**P<0.0001 between controls versus GOLD 2
†P<0.0001 between GOLD 1 versus GOLD 2

The overall distribution of the emphysematous tissue destruction (Lm) for all three groups was similar [Figure 3 & 4]. The majority of the Lm measurements fall between the 250-300 μm

range [Figure 3] and, although lower, is comparable to what was found in controls using microCT (57). Although the control distribution seems to have a number of measurements distributed in the 200-400 µm range (left shifted versus the overall distribution) the GOLD 1 and GOLD 2 groups have a similar distribution pattern [Figure 4] as the overall distribution [Figure 3].

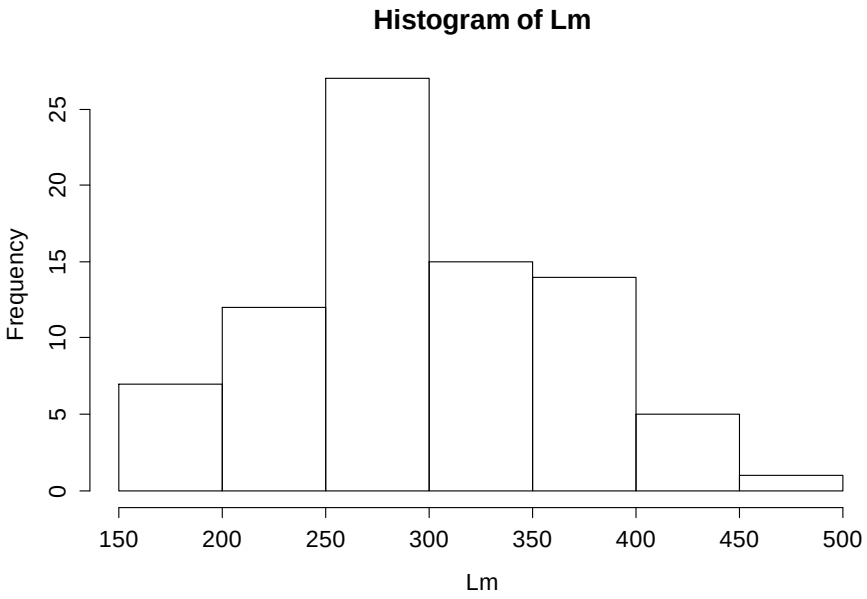


Figure 3: Overall Distribution of Lm in All Groups. The x-axis represents the range of Lm that was measured while the y-axis represents the number of observations for that respective Lm range. The range for each column is 50 um.

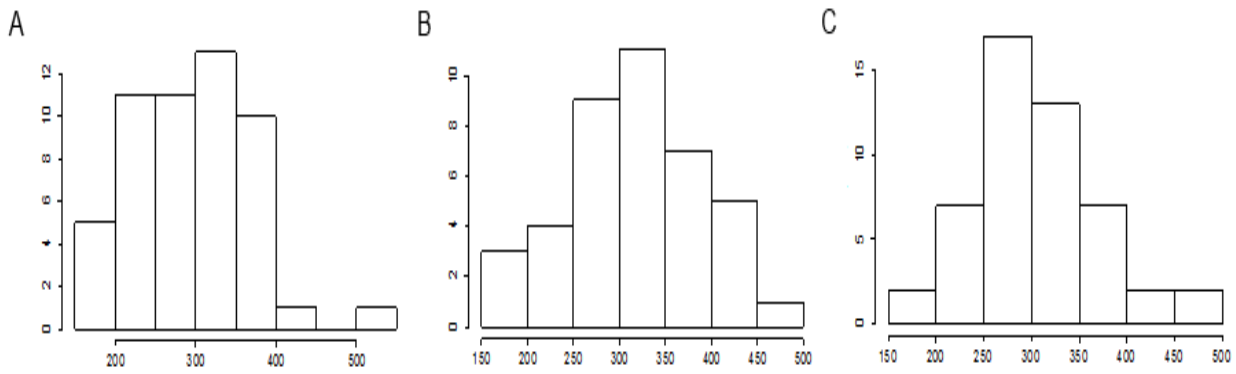


Figure 4: Histogram Breakdown of the Distribution of Lm amongst the Different Groups. The x-axis represents the range of Lm that was measured while the y-axis represents the number of observations for that respective Lm range. The range for each column is 50 um. A) Histogram for the control group. B) Histogram for the GOLD 1 group. C) Histogram for the GOLD 2 group.

Using the measurements obtained from both the immunohistochemical stains and the Lm measurements, comparisons using linear mixed-effect models were made [Table 7 & Figure 5]. The strongest correlations were between markers for macrophages, CD4+ T cells, CD8+ T cells, B cells, Lm, total tissue percent, and elastin. Overall, out of the 8 total significant correlations, 5 of these correlations included CD4+ T-cells (62.5%).

Table 7: Significant Results from Quantitative Histology Measurements within the Alveolar Tissue

Comparison	Coefficient	T-Stat	P-Value	FDR
CD68 (Alv) ~ CD4 (Alv)	0.519	6.20	1.34×10^{-8}	1.47×10^{-7}
CD4 (Alv) ~ Bcell (Alv)	0.451	4.88	3.80×10^{-6}	3.48×10^{-5}
CD68 (Alv) ~ Bcell (Alv)	0.358	3.83	4.33×10^{-4}	3.41×10^{-3}
Lm ~ Total Percent Tissue	-4.117	-3.03	2.65×10^{-3}	0.016
CD4 (Alv) ~ CD8 (Alv)	0.283	2.88	4.55×10^{-3}	0.025
CD68.Alv ~ Elastin (Alv)	-0.514	-2.68	8.67×10^{-3}	0.043
CD4 (Alv) ~ Lm	0.000	2.31	2.05×10^{-2}	0.087
CD4.Alv ~ Elastin (Alv)	-0.475	-2.34	2.01×10^{-2}	0.087

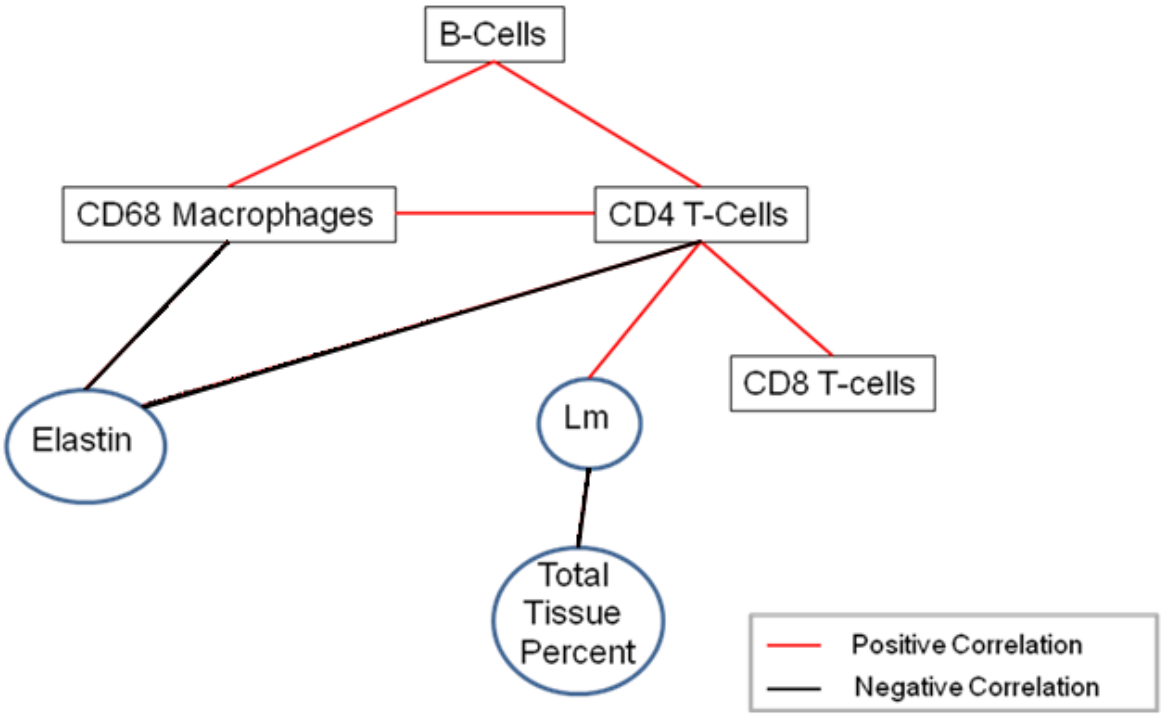


Figure 5: Summary Network of Significant Quantitative Histology Results. Structural components are represented by circles while the cellular components are represented by rectangles.

With respect to total 16S load there was a trend for a smaller total bacterial load in the GOLD 2 group versus the control and GOLD 1 group [Figure 6]. However, analysis with ANOVA found that there was no statistically significant difference between all three groups ($P > 0.05$). The overall bacterial load is similar and consistent with what has previously been reported (101,109).

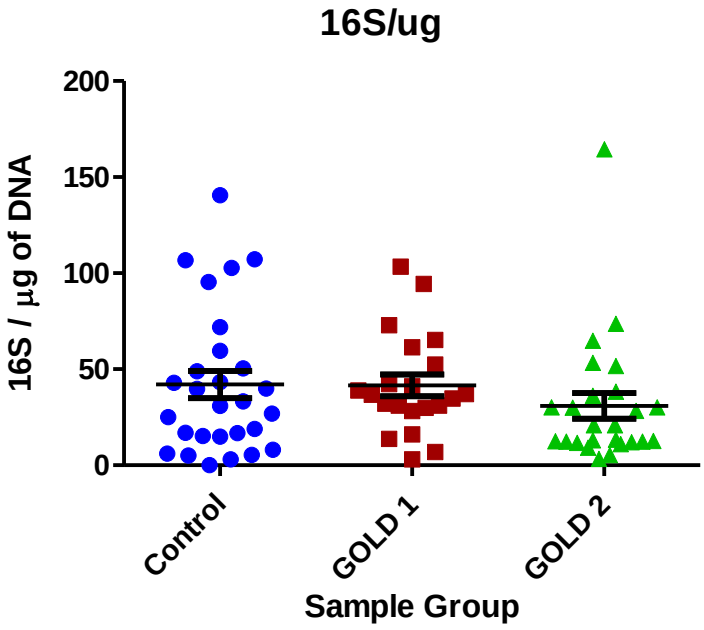


Figure 6: Overall total 16S Bacterial Loads of the Different Samples by COPD Grade. Total bacterial load was calculated by normalizing to the total DNA concentration of each sample used to the measured 16S from qPCR.

There was no difference in the alpha diversity of the bacterial community composition between the three groups (P-value > 0.05) [Table 8]. Further, there was no difference between Shannon diversity measures, evenness, or species richness (P-value > 0.05) between the three groups [Table 8]. However, there was a trend for increased species richness in the GOLD 1 and 2 groups versus controls. There was also a trend for decreased Shannon diversity and evenness between GOLD 2 versus GOLD 1 and controls.

When the alpha diversity was broken down by relative position, which was based on the location of where the lung resection was performed for each patient, there was no significant decrease in Shannon diversity between the three groups [Figure 7] (P < 0.05) at the top of the lung. In this

analysis top refers to a lung resection at the apex of the lung, middle to a resection from the middle of the lung, and bottom for a resection at the base of the lung [Figure 7-9]. It should be noted that there was a trend for decreased Shannon diversity in the GOLD 2 group when compared to either the control or GOLD 1 group at all three lung positions (top, middle, and bottom). Further, when looking at the absolute change versus the average Shannon Diversity (individual Shannon Diversity – Average GOLD Group Shannon Diversity (value obtained from table 8)) there was no significant difference between the three groups ($P > 0.05$) [Figure 7D-F]. However, it should be noted that at all three positions (top, middle, and bottom) the COPD GOLD 2 group had trends for increased absolute difference from the overall average Shannon diversity. In other words, there was a trend for a larger difference (positive or negative) from the average in the GOLD 2 group when compared to either the control or GOLD 1 group.

Table 8: Overall Average of Alpha Diversity Measurements of All Groups. Data is reported as mean ± standard deviation.

	Shannon Diversity	**Evenness**	**Species Richness**
Control "At Risk"	2.91 ± 0.28	0.86 ± 0.03	30.39 ± 7.62
GOLD 1	2.89 ± 0.22	0.86 ± 0.03	30.10 ± 5.69
GOLD 2	2.77 ± 0.60	0.83 ± 0.12	29.64 ± 10.73

There was no significant difference between the three groups when broken down by position in the lung for evenness [Figure 8] ($P > 0.05$). Although there was a slight trend for a decrease in evenness in the COPD GOLD 2 group when compared to both the control and GOLD 1 group. The absolute change in evenness (individual sample evenness – average GOLD group evenness (obtained from table 8)) showed no significant difference between the three groups by lung

position ($P > 0.05$). However, there was a trend for the GOLD 2 group to have a larger absolute change in evenness versus either the control or GOLD 1 group.

There was no significant difference between the three groups at top, middle and bottom when analyzing the species richness [Figure 9] ($P > 0.05$). When the absolute change in species richness was investigated there was also no significant difference between the three groups and relative lung position (top, middle, and bottom) ($P > 0.05$). However, there was a trend in the GOLD 2 group to have slightly higher species richness than the control and GOLD 1 group at all lung positions.

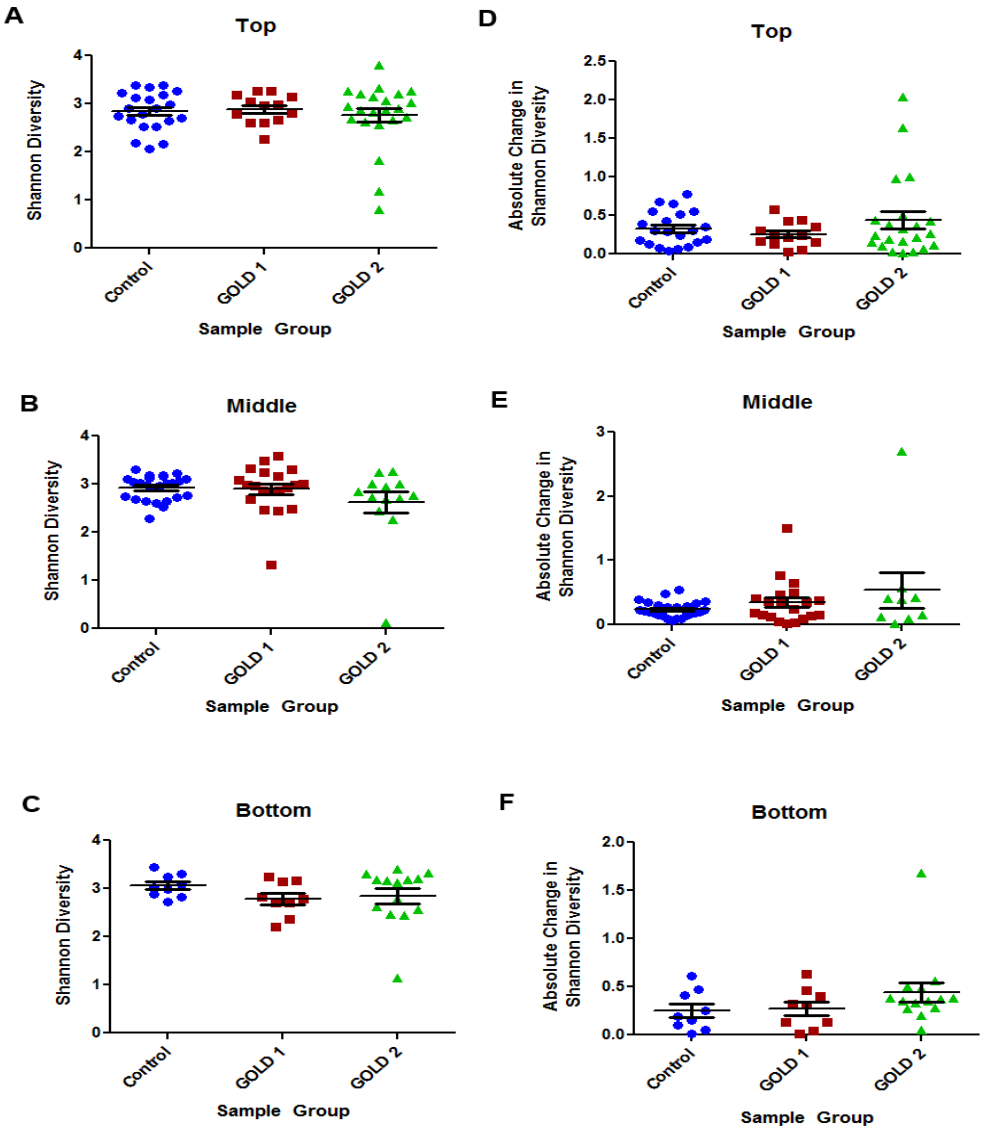


Figure 7: Breakdown of Shannon Diversity by Relative Position within the Lung Resection Sample. A) Shannon diversity in early COPD between control, GOLD 1, and GOLD 2 at the top of the lung. B) Shannon diversity in early COPD between control, GOLD 1, and GOLD 2 at the middle of the lung. C) Shannon diversity in early COPD between control, GOLD 1, and GOLD 2 at the bottom of the lung. D) Absolute change in Shannon diversity at the top of the lung. E) Average change in Shannon diversity at the middle of the lung. F) Average change in Shannon diversity at the bottom of the lung.

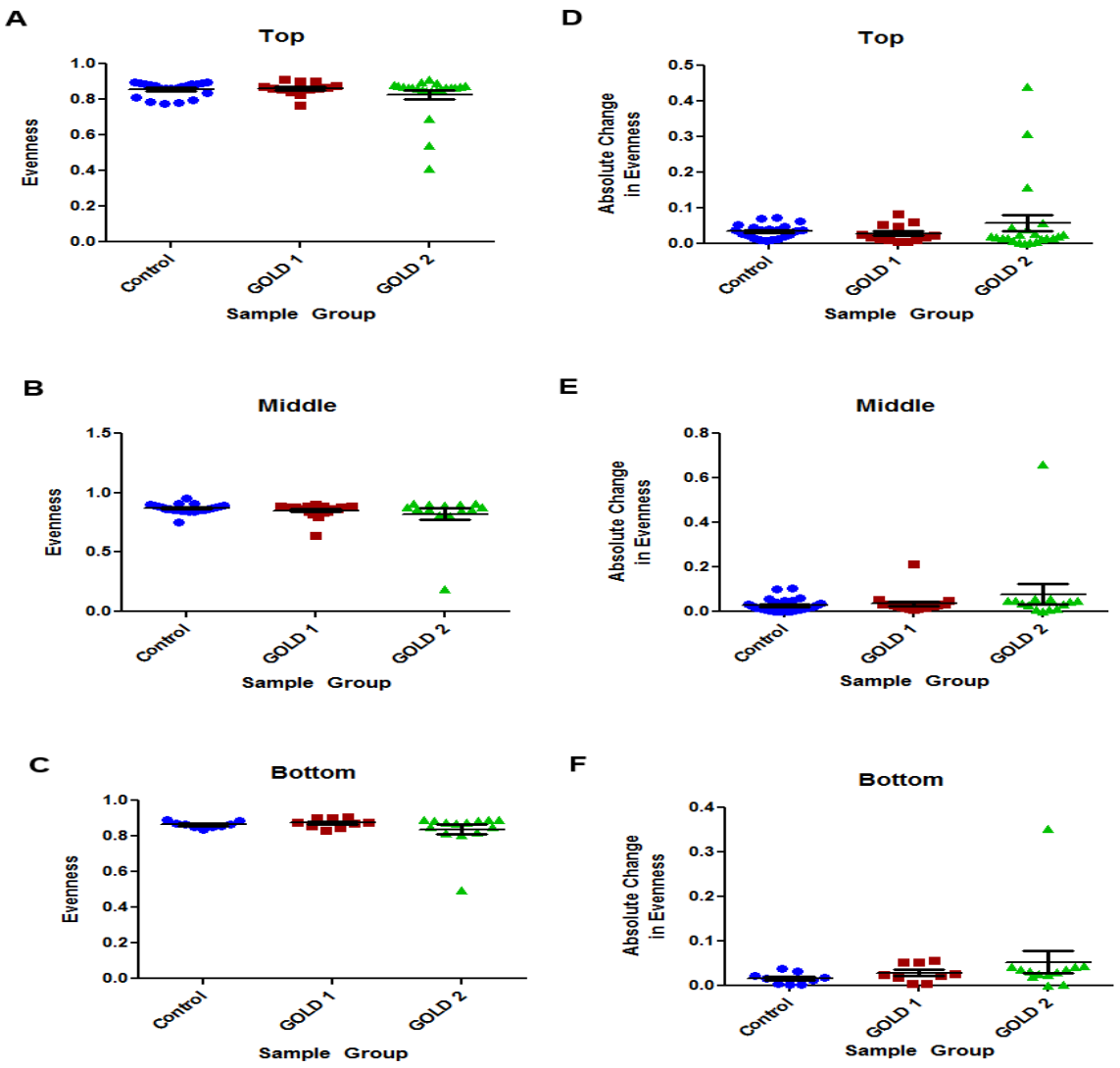


Figure 8: Breakdown of Evenness by Relative Position within the Lung Resection Sample. A) Evenness in early COPD between control, GOLD 1, and GOLD 2 at the top of the lung. B) Evenness in early COPD between control, GOLD 1, and GOLD 2 at the middle of the lung. C) Evenness in early COPD between control, GOLD 1, and GOLD 2 at the bottom of the lung. D) Absolute change from average evenness at the top of the lung. E) Absolute change from average evenness at the middle of the lung. F) Absolute change from average evenness at the bottom of the lung.

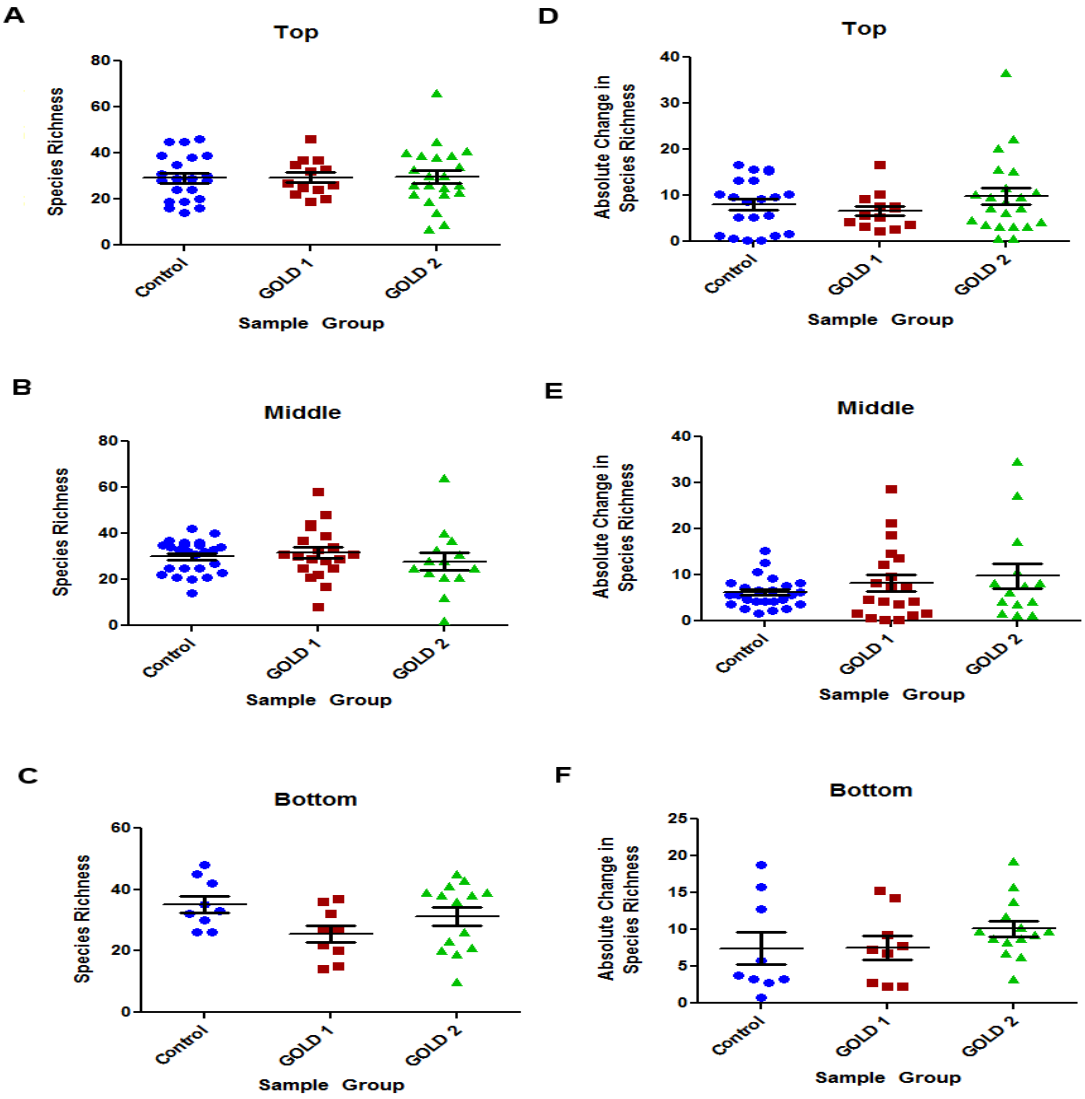


Figure 9: Breakdown of Species Richness by Relative Position within the Lung Resection Sample. A) Richness in early COPD between control, GOLD 1, and GOLD 2 at the top of the lung. B) Richness in early COPD between control, GOLD 1, and GOLD 2 at the middle of the lung. C) Richness in early COPD between control, GOLD 1, and GOLD 2 at the bottom of the lung. D) Absolute change from average richness at the top of the lung. E) Absolute change from average richness at the middle of the lung. F) Absolute change from average richness at the bottom of the lung.

There was a significant difference between the five groups tested, driven predominantly by the negative, extraction negative, and GOLD 2 groups (PERMANOVA < 0.05) [Figure 10]. This was done by using a post-hoc test on the first two components of the ordination. However, the GOLD 2 group was not significantly different than the negative control group (P>0.05) with respect to overall bacterial community composition. Interestingly a group of GOLD 2 samples seem to have quite a different community composition versus all the other samples [Figure 10].

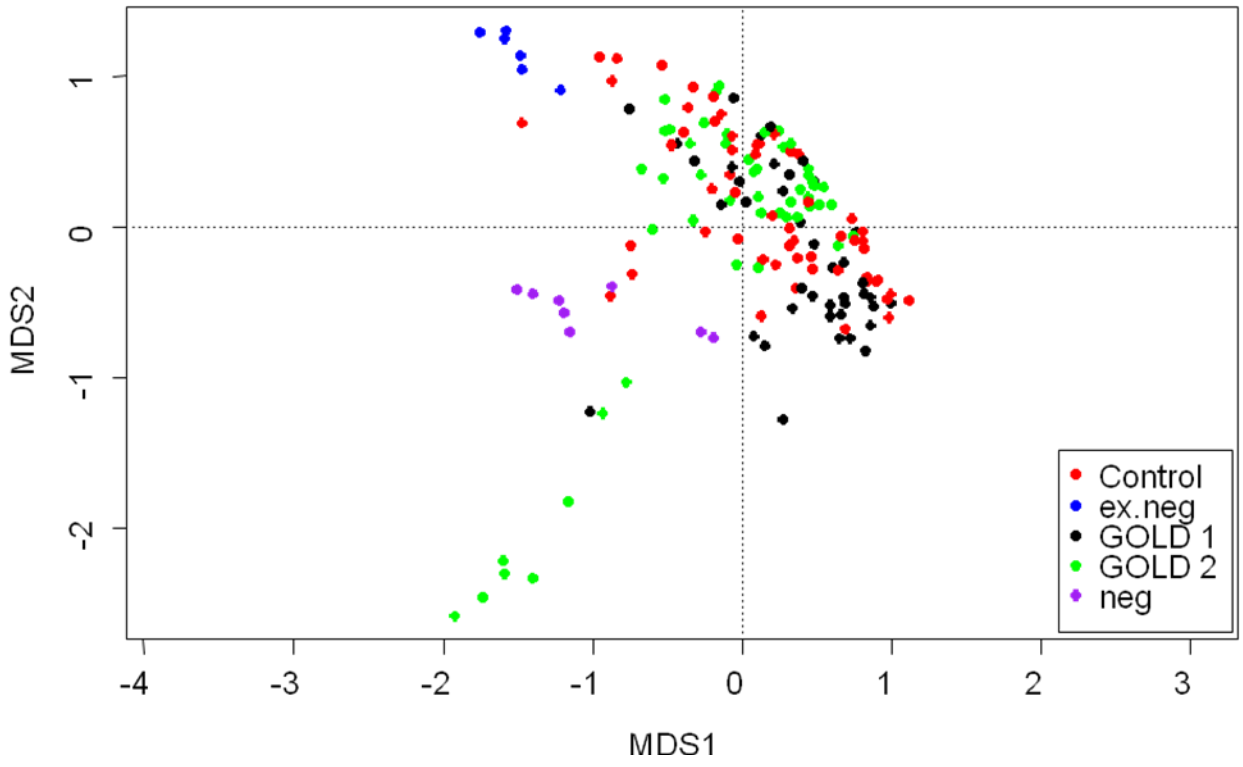


Figure 10: Non-Metric Multidimensional Scaling Analysis of the Bacterial Microbiome from the Mild and Moderate COPD Data Set. Extraction negatives are represented by ex.neg and negatives by neg.

When looking for whether or not the type of cancer had any relationship to the bacterial microbiome observed the initial PERMANOVA analysis showed a significant difference

between groups [Figure 11] ($P < 0.05$). However, when using the previously mentioned post-hoc analysis it was observed that these differences were driven by the extraction negative controls being different from some of the cancer groups (Large Cell Carcinoma (LCC), Non-Small Cell Carcinoma (NSCC), Adenocarcinoma (AC), Bronchioalveolar Carcinoma (BAC), and Carcinoma (CS)) as well as the extraction negative controls being significantly different from the negative controls [Figure 12] ($P < 0.05$).

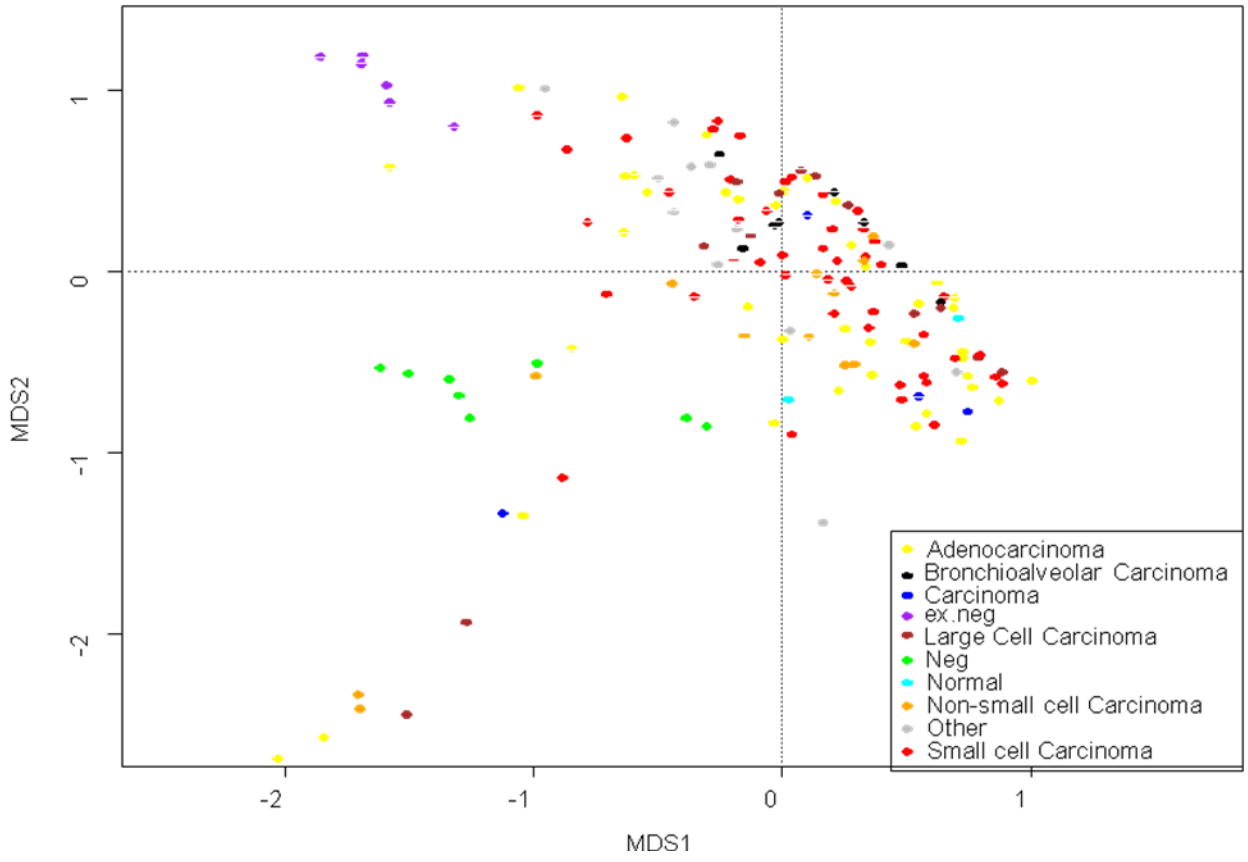


Figure 11: Non-Metric Multidimensional Scaling Analysis of the Mild and Moderate COPD Data Set Separated by Tumor Type.

	AC	BAC	CS	ex.neg	LCC	neg	Normal	NSCC	Other
AC	1.00000	-	-	-	-	-	-	-	-
BAC	1.00000	1.00000	-	-	-	-	-	-	-
CS	1.00000	1.00000	1.00000	-	-	-	-	-	-
ex.neg	0.01194	2.2e-05	0.00187	0.01756	-	-	-	-	-
LCC	1.00000	1.00000	1.00000	1.00000	7.5e-06	-	-	-	-
neg	1.00000	1.00000	1.00000	1.00000	1.7e-05	1.00000	-	-	-
Normal	1.00000	1.00000	1.00000	1.00000	0.55265	1.00000	1.00000	-	-
NSCC	1.00000	1.00000	1.00000	1.00000	1.4e-06	1.00000	1.00000	1.00000	-
Other	1.00000	1.00000	1.00000	1.00000	0.00630	1.00000	1.00000	1.00000	1.00000	-
SCC	1.00000	1.00000	1.00000	1.00000	0.00012	1.00000	1.00000	1.00000	0.90504	1.00000

P value adjustment method: bonferroni

Figure 12: Post-Hoc Test of the PERMANOVA for Mild and Moderate COPD which had a P-value < 0.05

Using Boruta feature selection with Random Forest analysis, 24 OTUs were found to be important in separating the different groups. The following heatmap shows that those with darker blue are found at a higher relative abundance and those found in red are at a lower relative abundance with white representing samples that did not contain those particular OTUs [Figure 13]. OTUs that aligned to *Flavobacteriaceae*, *Burkholderiales*, *Bacillaceae*, *Massilia Timonae*, *Sphingomonas*, *Fusobaccterium*, *Burkholderia Fungoru*, *Planomicrobium*, *Acenetobacter gyllenberi*, *Diaphorobacte*r, and *betaproteobacteria* were not found in the extraction negative controls. Another important observation is that one of the OTUs that aligned to *Massilia Timona* was only found in the GOLD 1 group and was not present in the other groups [Figure 13]. Additionally, an OTU that aligned with the Fusobacterium family was only present in the GOLD 2 group [Figure 13]. The differences between the controls, GOLD 1, and GOLD 2 groups showed that many of the OTUs were present in the control group but not in GOLD 1 or GOLD 2 (e.g. *Burkholderiales*, *Burkholderia Fungoru*, and *Planomicrobium*) suggesting that these 24 discriminative OTUs were able to separate between control, mild, and moderate COPD mostly

due to a generalized loss of these OTUs from the control and GOLD 1 groups [Figure 13]. Finally, there were a few OTUs that were present in control and GOLD 2 groups but not in the GOLD 1 group (e.g. *Ralstonia*, *Actinomycetales*, *Burkholderiaceae*, and *Burkholderia* 0X-0).

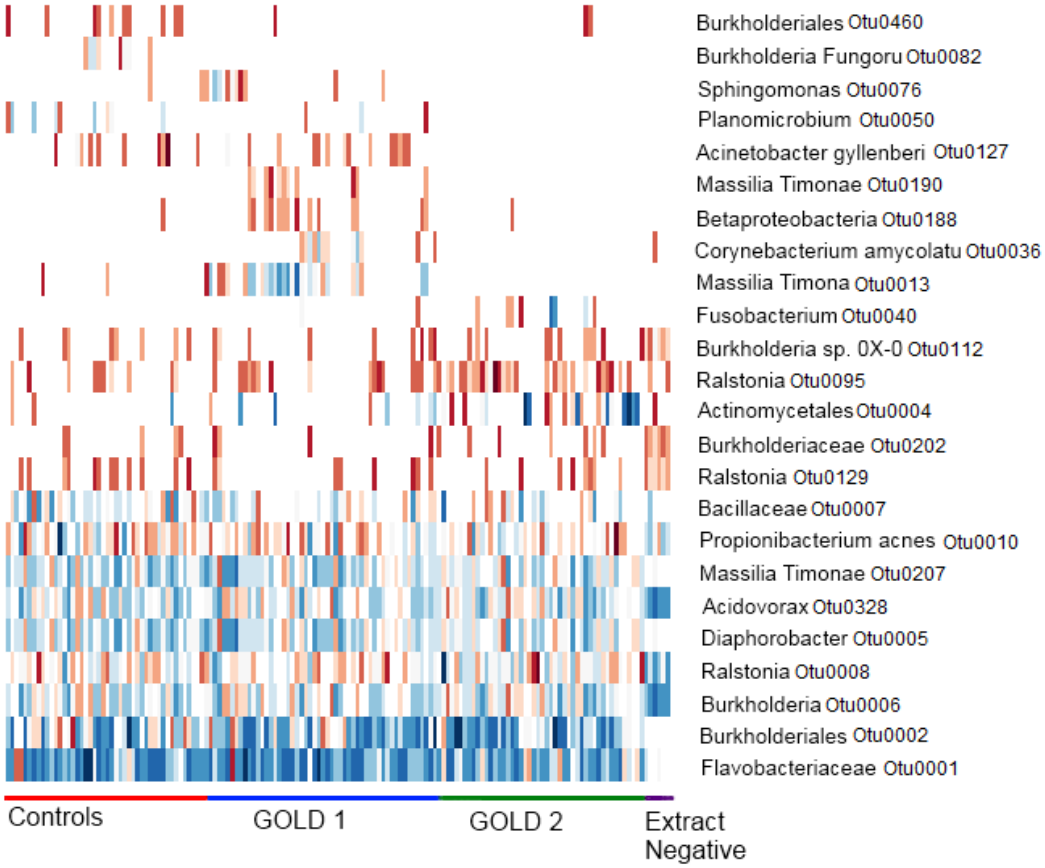


Figure 13: Heatmap of Boruta Picked Discriminative OTUs between the Different Groups. White represents not present, red represents low relative abundance, and blue represents high relative abundance.

There were no significant correlations between Lm and the microbiome in mild and moderate COPD. Using regularized canonical correlation analysis (RCCA) it was found that the

Proteobacteria phyla were negatively correlated with alveolar B-cells, CD4 T-cells, and macrophages. The *Bacteroidetes* phyla were positively correlated with alveolar macrophages and CD4+ T-cells and the *Firmicutes* phyla were positively correlated with total tissue percent [Figure 14 & 15]. As a brief aside in figure 15 the positive correlations between variables are when two measurements are closely clustered together (e.g. Bacteroidetes and CD4. Alv) while negative correlations can be visualized by two variables that are far apart (e.g. Proteobacteria and CD4.Alv). However, when the data was analyzed using a linear mixed-effects model, to account for multiple samples from the same individual, these correlations were no longer significant (FDR > 0.1).

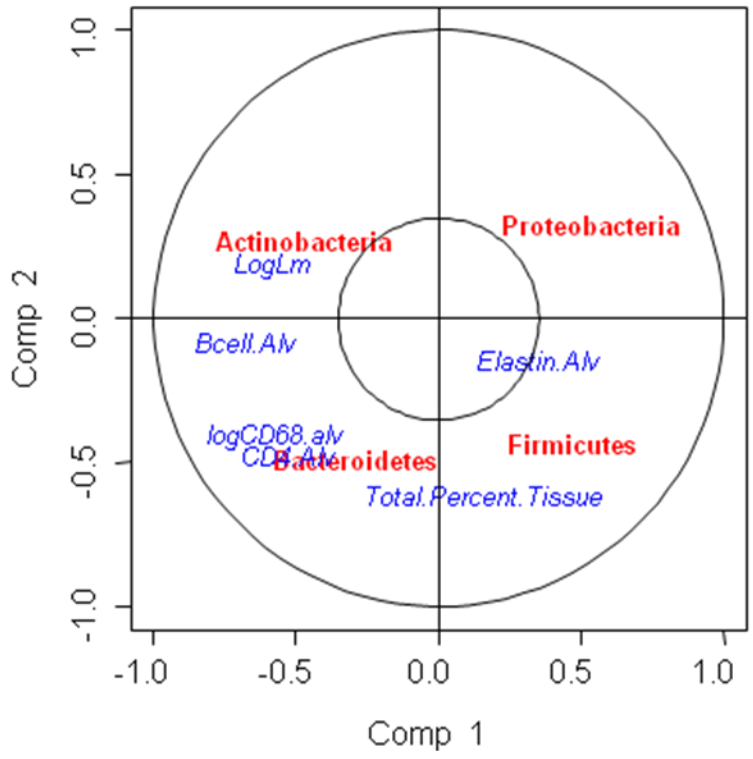


Figure 14: Graph of the Correlations between the Microbiome and Quantitative Histology in the Mild and Moderate COPD Data Set. Correlation cutoff was set an R of 0.35.

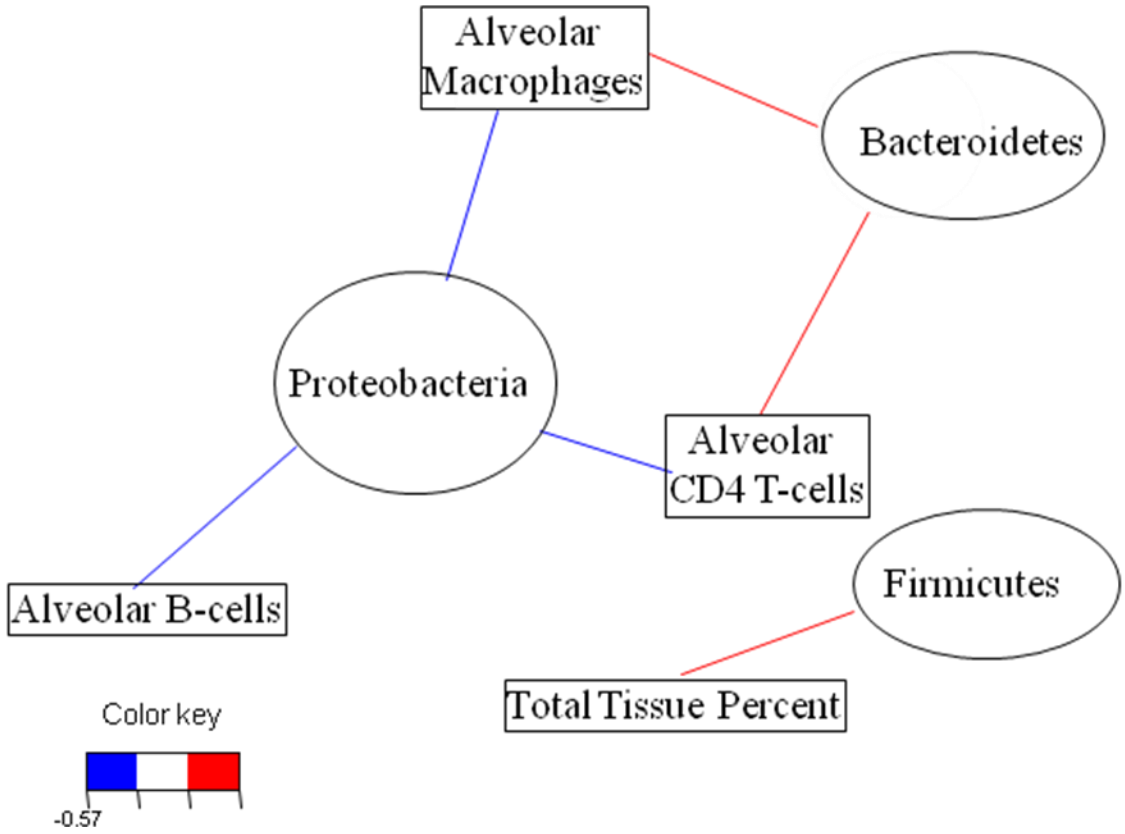


Figure 15: **Network of the strongest Correlations between Histology and the Microbiome in Mild and Moderate COPD Data Set**. The R value cutoff used was 0.50. Red represents positive correlations over the R value cutoff of 0.5 while blue represents the R value cutoff under -0.5 for the negative correlations. The lowest R value possible for the negative correlations was -0.57.

3.4 Discussion

Within areas of lung without appreciable emphysema, as measured by Lm, there were significant amounts of inflammation. The major inflammatory cells that were involved were macrophages, B cells, and CD4+ T-cells. In particular, the CD4+ T-cells were a central hub positively correlated with Lm, CD68+ macrophages, and B cells. It is possible that certain bacteria that are

either lost or gained in the bacterial lung tissue microbiome during mild and moderate COPD could provide the antigenic targets for these specific inflammatory and immune cells.

One notable absence in this data set versus others is the correlation between CD8+ T-cells versus specific disease markers of COPD severity. Previous research has shown that CD8+ T-cells are increased in COPD versus smokers (154) and can undergo reduced apoptosis (155). In contrast this data set shows that CD8+ T-cells are not correlated with Lm. There are a few possible explanations to this. First, although not correlated directly to Lm the CD8+ T-cells are positively correlated with CD4+ T-cells [Figure 5]. So it could be possible that the CD8+ T-cells do have an impact on severity but it is an indirect one. In fact a very recent study has shown that CD8+ T-cells can indeed help activate CD4+ T-cells (156) and this might be what is going on in mild and moderate COPD. Second, this research was also one of the first to investigate Lm specifically in mild and moderate COPD and how it could relate to various inflammatory cells. Thus it cannot be discounted that CD8+ T-cells simply do not correlate with Lm. Newer published abstract data from the Hogg lab would support this position (138,139).

There was no difference in total bacterial 16S load between controls, GOLD 1, and GOLD 2 groups. However, there was a difference in the bacterial community composition between GOLD 2 versus GOLD 1 and control groups. These changes were not due to any particular increase or decrease in Shannon diversity. Additionally, cancer diagnosis did not make a significant contribution to the differences observed in the GOLD 2 group. A total of 24 specific bacterial OTUs were identified to be important in discriminating the three groups from each

other and by using RCCA it was found that certain phyla had correlations with particular quantitative histological measures of inflammation.

This data suggests that, at the earliest stages of the disease, an active, robust, localized inflammatory response is occurring. This response could be the predominant driver of the emphysematous destruction seen in later more severe disease. Further, this inflammation involves macrophages, CD4+ T-cells, and B cells suggesting that an active adaptive immune response is present even at the earliest grades of the disease. The correlation between CD4+ T-cells and Lm suggest that this adaptive immune response is directly related to emphysematous tissue destruction. This link between CD4+ T-cells and disease progression has been shown previously (53,54) but never in mild and moderate disease before any emphysematous tissue destruction is clearly noticeable (57,58).

Although there are some limitations to the Lm measurement such should not drastically change the findings. Some limitations to the Lm measurement include shrinkage from the fixation used as well as the OCT inflation of the lung tissue. Fixation shrinkage due to the use of formalin is well documented in the literature (157–159) and could artificially lower the Lm measurements such that they are artificially below the 95% confidence interval of 495 μm observed for Lm seen in controls from a previous study (57). The second limitation is the use of OCT for inflation of lung tissue. Although the inflation was done at a constant pressure, OCT is rather sticky and can get stuck while being perfused through the lung. Thus certain areas will encounter incomplete inflation and artificially small Lm values. In order to correct for this, sections that looked artificially compressed were excluded from the analysis. Further studies,

whose aims are to control for these factors, need to be completed to confirm these results and findings in mild and moderate COPD.

There was no significant difference between the 16S bacterial load in all three groups and this is consistent with previous research (101,109). There was a slight overall decrease in the total bacterial 16S load in the GOLD 2 group and this could suggest that there was a potential loss of a small number of bacterial species without any corresponding increase in any of the other bacterial species present. From previous data, COPD GOLD 4 total bacterial 16S load is similar to that of both smokers without airflow limitation and non-smokers with normal lung function (109). Thus if there is a loss of bacterial species that manifest in decreased total bacterial 16S load this might eventually rebound in later disease, perhaps by a replacement with those bacterial species that are able to survive and thrive in that particular environment. This can be partially supported by research done in BAL and bronchial brushings of moderate COPD where an overall decrease in diversity compared to controls was observed (107) yet no such difference was observed in later disease (109).

Although a significant difference was found between the GOLD 2 versus control or GOLD 1 group using NMDS analysis, no such differences were found in the alpha diversity along with its different components (Shannon Diversity, evenness, richness). This information could indicate that the bacterial community changes are small and involve particular OTUs rather than large scale community changes. These changes do not manifest in significant differences in evenness or in changes to the species richness. The loss of OTUs and potentially bacterial diversity would

be consistent with other chronic inflammatory diseases in which the bacterial microbiome plays an important role (105,106).

One of the potential limitations of this data set was that it was obtained as a population of convenience and every individual in the study had a co-variable of a cancer diagnosis. Analysis showed that there was no difference by cancer diagnosis except that most of the groups were significantly different from both the extraction negative and negative controls. Thus cancer in this population could be ruled out as a possible confounder of this patient population and data set with respect to the bacterial microbiome.

Many OTUs are lost in both GOLD 1 and GOLD 2 versus the controls [Figure 13]. Yet even with this loss there are specific OTUs that are only seen in either GOLD 1 (e.g. *Massilia Timonae*) or GOLD 2 (e.g. *Fusobacterium*). What is observed could possibly be an early disturbance of the bacterial microbiome, with a small number of OTUs being lost and replaced as the bacterial lung tissue microbiome moves towards what is observed in GOLD 4 grade disease. These specific OTUs that are lost may have specific anti-inflammatory roles, interactions, or just may be bystanders that could no longer adapt to live in the toxic environment created by the combination of inflammation and remodeling. A recent study by Salter, et al. (160) suggests that a large number of the bacteria identified in the mild and moderate COPD microbiome could be from contamination. However, some of the species and genera identified by this study (Pseudomonas, Ralstonia, Streptococcus, etc.) have been previously identified as either specific to the lung or important in discriminating between control and disease (101,103). Based on this information caution does need to be exercised when interpreting the results that were obtained in

this study. It is likely that Massilia, Sphingomonas, and certain Burkholderia genera are environmental contaminants. In contrast, Ralstonia may be a true signal based on the previous literature. Overall, there may be a few OTUs that represent genuine bacteria from the lung tissue microbiome.

Another potential limitation was that many OTUs were not represented in every sample, with a vast majority of the OTUs only present in less than 50%. This makes statistical analysis using linear models difficult since the many missing values tend to lead to non-normal distributions, making analysis skewed and easily influenced by outliers. With this in mind phyla were utilized due to their relatively continuous distribution throughout every sample. However, this leads to a new limitation in that different bacteria within one specific phylum will not have the same role as others. Using phyla for the analysis allows for the use of sophisticated linear models in the analysis but in doing so we have to make broad generalizations that do not necessarily hold true for each bacterial species within that phylum.

Using RCCA it was possible to show that certain phyla were correlated with specific inflammatory and immune cells. The phyla most closely correlated with Lm were the *Actinobacteria* [Figure 14]. Another interesting observation was that the *Firmicutes* phylum was positively correlated with total tissue percent [Figure 14]. Additionally, Lm and total tissue percent are negatively correlated [Figure 14]. This provides evidence that the phyla, *Firmicutes* and *Actinobacteria* are either involved in the loss of alveolar tissue, protective against tissue loss, or are bystanders to this loss and are reduced simply due to the fact that they are losing particular places in which they can thrive. Since no active inflammatory response is correlated with either

of these two phyla it would suggest that the more plausible mechanism is that of a bystander. These two phyla could simply decrease as a consequence of the inflammation and remodeling that occurs.

Increasing the R value cutoff to 0.5 removes the *Actinobacteria* and Lm correlation, but nicely shows a dynamic interplay between *Proteobacteria* and *Bacteroidetes* phyla and their relationship with CD4+ T-cells and CD68+ macrophages within the alveolar tissue [Figure 15]. What is interesting is that certain bacteria within the *Proteobacteria* and *Bacteroidetes* phyla seem to be antagonizing of each other. *Proteobacteria* are negatively correlated with B cells, CD4+ T-cells, and CD68+ macrophages. This would suggest that specific or multiple bacterial species within this phylum could be targeted by an adaptive immune response. Alternatively, *Bacteroidetes* is positively correlated with both CD68+ macrophages and CD4+ T-cells. This would suggest that as particular bacterial species within the *Proteobacteria* phyla are destroyed more bacteria from the *Bacteroidetes* phylum grow and potentially fill the void left by bacteria from the *Proteobacteria* phyla. This analysis suggests that even in the earliest grades of disease there could be an adaptive immune response to bacteria within specific phyla in COPD. Collectively, this data suggests that in mild and moderate disease, before Lm increases significantly, there are close correlations between CD4+ T-cells, B-cells, and CD68+ macrophages occurring. While this occurs specific OTUs are either decreased or increased within the bacterial lung tissue microbiome. Ultimately these small changes result in a significant difference between the GOLD 2 bacterial community composition and those of the GOLD 1 and control groups. Overall comparison of the histological measurements with the bacterial lung tissue microbiome suggests that the bacteria within the *Proteobacteria* and

Bacteroidetes phyla could be potentially important players in the severity of COPD due to their close correlations with important cells that have previously been correlated with disease severity (53,61,161).

From this chapter it can be shown that the bacterial microbiome within the lung tissue changes from controls to COPD GOLD grade 2, although no direct evidence of correlations of specific OTUs could be found to structural changes or inflammatory changes within the lung tissue. There were interesting data showing that changing the balance of different phyla could potentially have an impact on inflammatory cell infiltration into the lung tissue. Such suggests that a changing bacterial microbiome could influence certain cellular components involved with COPD pathogenesis. In the next chapter the host response to the bacterial microbiome is explored in more detail. In addition, I try to find specific OTUs within this bacterial microbiome that may be important to disease progression.

Chapter 4: The Host Response to the Bacterial Microbiome in COPD[2]

4.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive, debilitating lung disease with multiple co-morbidities that affects more than 200 million people worldwide and is responsible for approximately 3 million deaths each year (3). Although the pathogenesis of small airways obstruction and emphysematous destruction responsible for the progressive airflow limitation in COPD has been associated with the host innate and adaptive inflammatory immune response (53,137,162), the antigens that drive this response remain poorly understood. The British Hypothesis, that smoking compromised the host response to allow colonization and infection of the lower respiratory tract by organisms that caused chronic bronchitis and the decline in the forced expiratory volume, was rejected based on a prospective longitudinal study conducted by Fletcher and associates (163). This study showed that many people with chronic bronchitis never developed airflow limitations and that many others developed severe airway obstruction in the absence of chronic bronchitis (163,164). Sethi, Murphy and their colleagues reawakened interest in the possible role of bacteria in the pathogenesis of COPD by showing that acute exacerbations of COPD were commonly associated with the emergence of new bacterial strains that could be isolated from the sputum and protected bronchial brushings (142). Moreover, Wedzicha and her associates extended these observations by showing that frequent

[2] This section has been published in the American Journal of Respiratory and Critical Care Medicine.
Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, Erb-Downward JR, Huffnagle GB, Hayashi S, Elliott WM, Cooper JD, Sin DD, Lenburg ME, Spira A, Mohn WW, Hogg JC. The host response to the lung microbiome in chronic obstructive pulmonary disease. 2015 (Accepted)

exacerbations of COPD within the same individual are associated with an accelerated rate of decline in lung function leading to COPD (77).

The application of culture independent techniques to the identification and community analysis of bacteria led to the discovery that the human gastrointestinal and genitourinary tracts, as well as in the skin, mouth and upper airways host relatively large and complex microbiomes that live commensally within the host (165–168). In contrast the long held view that the lung was sterile below the larynx persisted until Hilty and associates (107) used these techniques to challenge this hypothesis by analyzing the microbiome in bronchial brushings and washings from human lungs. Their new data suggested that lower airways of patients with asthma and COPD contained a microbiome that became less diverse and was associated with the emergence of potential pathogens (107). Although these results were criticized as artifact produced by contamination of the bronchial brushings and washing as they passed through the upper airways, this criticism was refuted by Erb-Downward, et al (101) and Sze , et al. (109) in studies that demonstrated a human lung microbiome in samples obtained by either brushing the airways of explanted lungs where the upper airways were absent (101) or rapidly freezing the explanted lung solid to allow peripheral lung samples to be removed without disturbing the central airways (109). The present report extends these observations by examining the microbiome in relation to emphysematous destruction of the lung gas-exchange surface and providing preliminary evidence that this destruction is associated with the development of a host immune response to this microbiome. Some of the results of these studies have been previously reported in the form of an abstract (169–172).

4.2 Methods

4.2.1 Consent

Informed consent was obtained either directly from the patients being treated for very severe COPD by lung transplantation, or from the next of kin of organ donors who agreed that the lungs could be released to serve as controls when considered unsuitable for transplantation. The conditions under which consent was obtained were approved by the appropriate committees at each of the participating institutions (57,173) and the shipment of specimens between institutions was compliant with the US Health Insurance Portability and Accountability Act.

4.2.2 Specimen Preparation

Specimen preparation has been described in detail in our previous publications (19,57,173) and in the online supplement. Briefly 5 explanted lungs from patients with GOLD 4 COPD and the 4 donor (control) lungs were fully inflated with air to 30 cm trans pulmonary pressure (P_L) and then deflated and held at a P_L of 10 cm H_2O while frozen solid in liquid nitrogen vapor. These lung specimens were kept frozen on dry ice while a volumetric multi-detector computed tomography (MDCT) scan was obtained and while the specimen was cut into contiguous 2cm thick transverse slices from lung apex to its base. A cluster of 4 cores of lung tissue was removed from each slice for each of the investigations outlined below.

4.2.3 Microbiome Analysis

The pipeline for protocol 1 is fully described by Schloss et al. (174,175) and was developed at the University of Michigan based on touchdown PCR amplification of the V3-V5 region of the bacterial 16S rRNA gene with pyrotag sequencing of the amplified DNA at the University of Michigan Microbiome Sequencing Facility using a low-biomass protocol (101,103,176,177). The pipeline for protocol 2 used to analyze the bacterial 16S ribosomal DNA is fully described in our previous publications (53,57,109,173) and the online supplement. Protocol 2 was developed in Vancouver and based on nested PCR amplification of the V1 to V3 region of the bacterial 16S rRNA gene and pyrotag sequencing of the amplified DNA by Genome Quebec (109). It was used as an independent method to confirm microbiome results obtained from protocol 1 [Figure 16].

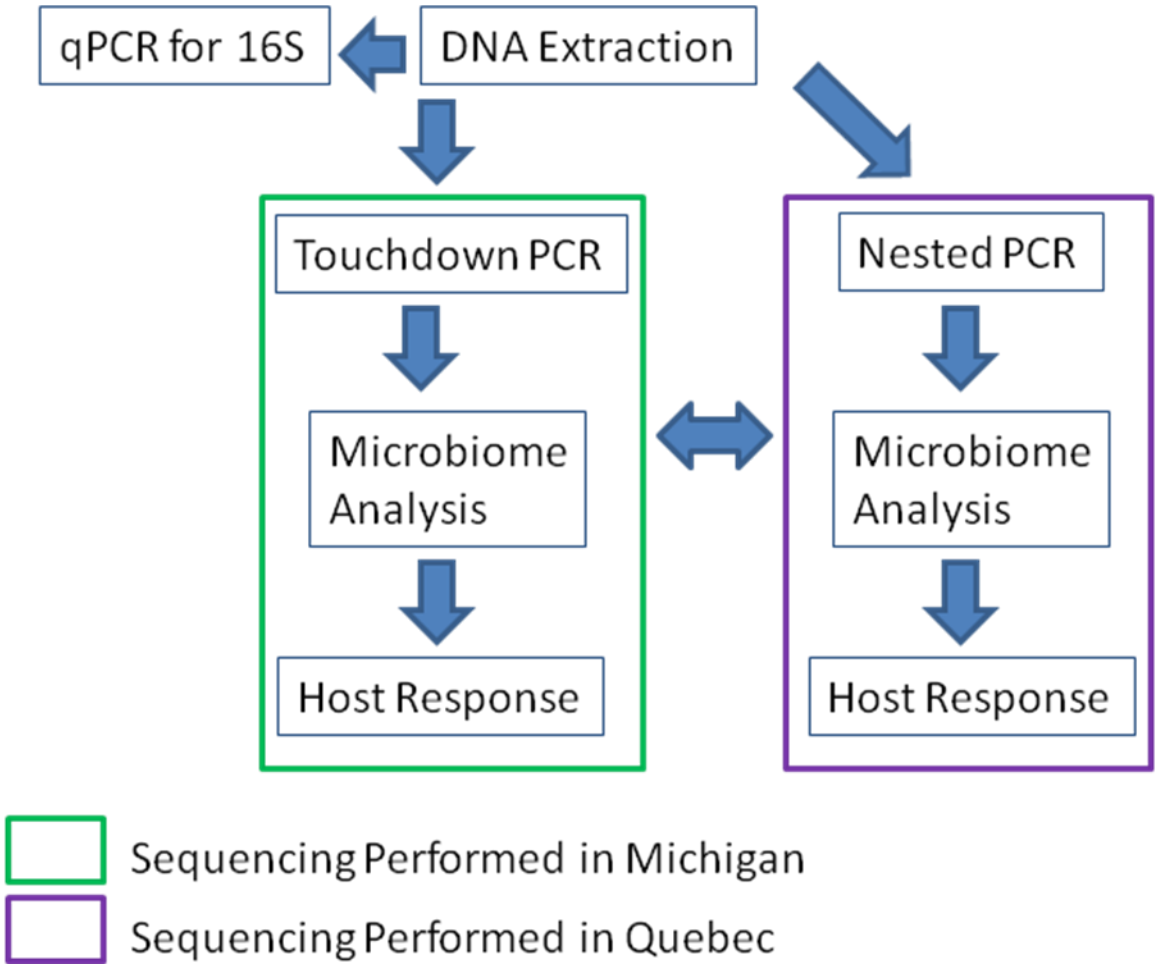


Figure 16: Overall Workflow of the Bacterial Microbiome Analysis.

4.2.4 Microbial Diversity

Microbial diversity was assessed using:

$$H = E_H \times \ln S \quad \text{(Equation 1)}$$

where H is the Shannon diversity index, E_H represents the evenness of the community of OTUs in the sample and lnS represents the natural log of OTU richness (or numbers of different OTUs). Differences between the bacterial community composition in control and COPD lung

samples were visualized using principle components analysis (PCA) of pair-wise Bray-Curtis dissimilarities and tested with permutational multivariate analysis of variance (PERMANOVA) (148).

4.2.5 Emphysematous Destruction

Emphysematous destruction was assessed by measuring the alveolar surface area (SA) of each lung sample:

$$SA = 4 \times V / Lm \quad \text{(Equation 2)}$$

where SA is the internal surface area of the core of lung tissue removed at each of the sampled sites, V is the total volume of lung in the tissue core removed from the lung and Lm is the mean linear intercept.

4.2.6 Immune Cell Infiltration

The infiltration of inflammatory immune cells into the tissue was estimated by point counting the volume fraction (Vv) of the bronchiolar and alveolar tissues occupied by polymorphonuclear leukocytes (PMN), macrophages, CD4+, CD8+ and B lymphocytes on appropriately stained histological sections from companion cores of tissue to those examined by microCT in the lungs from patients treated by lung transplantation and their controls.

4.2.7 Host Gene Expression

Detailed methods for gene expression profiling can be found in the online supplement (Section 4.4) and in a previously published manuscript (19). These gene expression data are available through the Gene Expression Omnibus (GEO) under the accession GSE27597. In total four different core tissue samples were used, one for each part of the analysis.

4.2.8 Statistics

A linear mixed-effects model was used to compare OTU richness to emphysematous destruction assessed from measurements of the lung surface area, as well as the host response to this tissue destruction. These were obtained by Vv of the tissue occupied by inflammatory immune cells or gene expression profiling studies conducted on the RNA isolated from histological sections cut in close proximity to those examined by histology. The linear mixed- effects model allowed correction for the effect of lung height and position of samples within each lung slice (19). Gene expression pathways were further analyzed using DAVID (178).

Only the phyla and families that achieved significant correlations with at least one of Vv or microCT measurements were compared to host gene expression. If a phyla or family was undetected in more than 30% of the samples the data were converted to a categorical variable (positive or negative) and then analyzed using the linear mixed-effects model. To identify the OTUs that were most likely driving the correlations with phyla, the data were separated based on the average value of the host measurement of interest and a high and low group were created. If

an OTU was significantly different between these two groups and matched the direction of correlation found in the phyla analysis, it was considered a potentially important OTU. Additionally, OTUs identified by Boruta feature selection (152) after Random Forest analysis, as discriminative for control and GOLD 4, were also analyzed using linear mixed-effect models and compared to microCT, quantitative histology, and gene expression data. Gene Set Enrichment Analysis (GSEA) was used to compare similarity in the overall gene expression data sets. Further details on the full data analysis are provided in the online supplement.

4.3 Results

Table 9 and Table 13-14 of the online supplement summarize the data concerning age, gender, smoking history, lung function, number of tissue samples used for each analysis, and the number of reads per sample on all the subjects in this study.

Table 9: Demographics Data for Patients Analyzed

	Controls (n=4)	**GOLD 4** (n=5)
Age	53.8 ± 4.3	60.0 ± 1.6
Sex (M:F:Unknown)	4:0:0	3:2:0
FEV1/FVC	N/A	0.31 ± 0.07
FEV1 % Predicted	N/A	17.89 ± 5.47
Samples / Individual (n)	8 (3), 5(1)	8(5)

Microbial diversity as measured by OTU richness declined as emphysematous destruction increased [Figure 17] and that there was a linear correlation ($R^2 = 0.27$) between OTU richness and alveolar surface area. This was confirmed after applying a second independent protocol to assess the microbiome [Figure 19]. Furthermore, the PCA showed differences in the bacterial communities between GOLD 4 lung tissue samples and the control lungs (Figure 17B, P = 0.001) based on PERMANOVA (179). This difference was also found using the alternative protocol [Figure 19B] ($p< 0.01$). Although there was a trend for Shannon diversity to be lower in samples of lung from patients with GOLD 4 COPD this difference only became statistically significant ($P<0.05$) in samples from position 6 [Figure 17C] (1=apex 12=base). The difference observed between the controls and GOLD 4 were 3.40 ± 0.24 vs. 2.25 ± 0.69, (mean ± SD) and the negative water controls 1.6 ± 0.1.

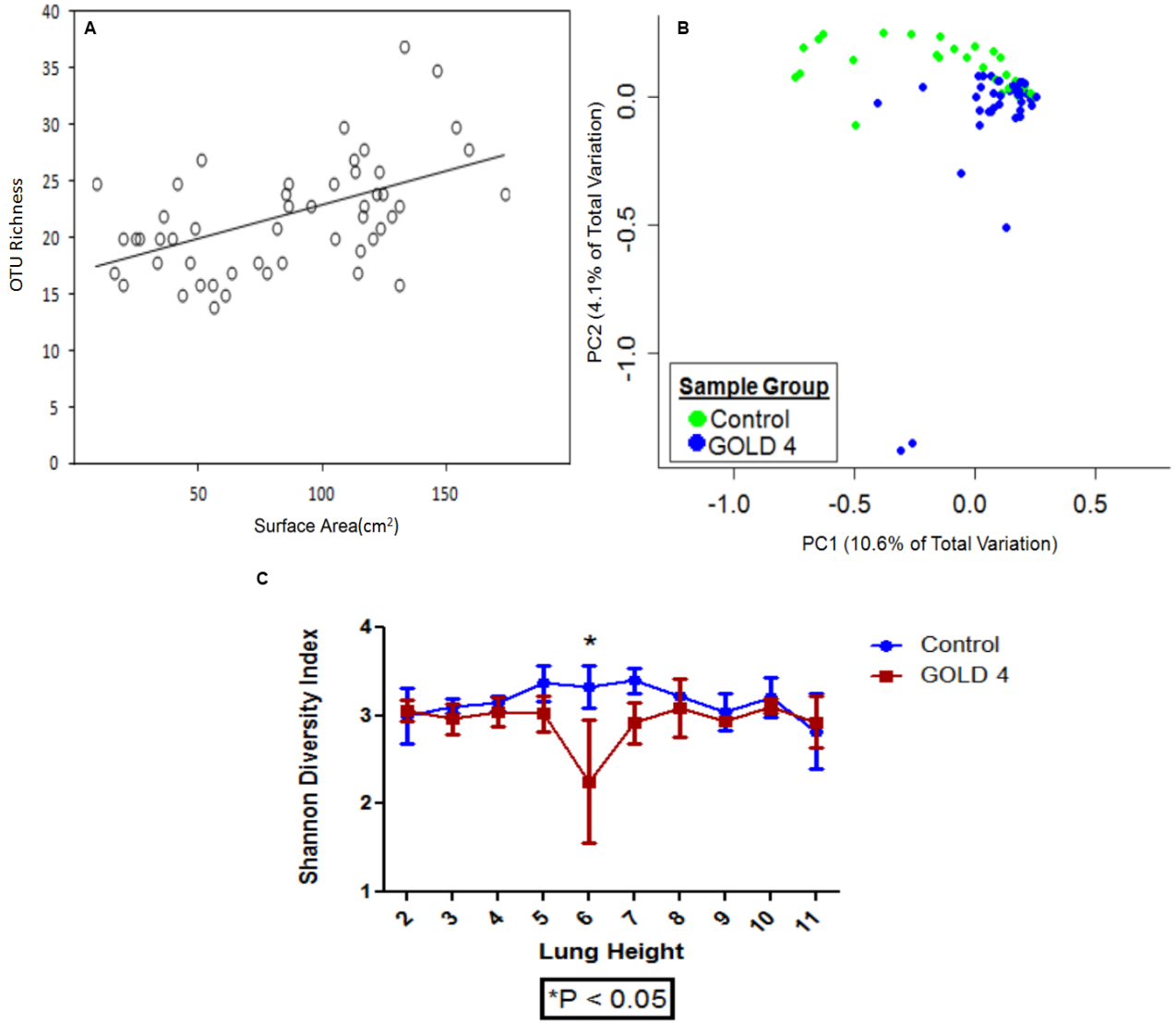
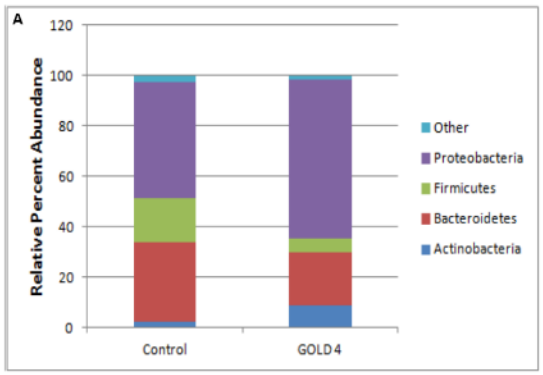


Figure 17: Protocol 1 or Touchdown Approach with V3-V5 primers. **OTU richness as a function of alveolar surface area (A), Ordination of samples based on Bray-Curtis dissimilarity of microbiomes (B). Shannon Diversity versus lung height between control and GOLD 4 (C)**. A) Alveolar surface area values and OTU richness determined from spatially adjacent cores (R^2 = 0.27, P<0.05). B) Dissimilarity was calculated using the same approach as graph C. The two groups were significantly different (PERMANOVA; pseudo-F=6.58; P=0.001). C) Lower lung height values represent lung tissue taken closer to the apex while higher lung height values represents lung tissue taken closer to the base. There was a significant difference between control and GOLD 4 (P < 0.05) at the relative middle of the lung

Relative abundances of bacterial phyla differed ($P < 0.05$) between GOLD 4 and control lungs [Figure 18A]. Based on Bonferroni post hoc testing the expansion of the Proteobacteria phylum was the most significant driver of the difference between control and GOLD 4 ($P < 0.05$). Overall, the largest differences between the two groups was seen between the Proteobacteria (controls: 46 ± 16%, GOLD 4: 66.0 ± 1.6%), Firmicutes (controls: 17.7 ± 19.6 %, GOLD 4: 7.04 ± 0.87%), and Bacteroidetes (controls: 31.7 ± 11.3 %, GOLD 4: 21.1 ± 4.1 %). The Proteobacteria, *Haemophilus influenzae*, was among the 10 OTU's that were important for discriminating between the control and GOLD 4 bacterial microbiome [Figure 18B] according to Boruta feature selection with the Random Forest analysis. Although the majority of these bacterial species decreased in abundance in GOLD 4 COPD lung tissue, a notable exception was *Elizabethkingia meningoseptica* [Figure 18C]. The similarity and differences between the two methods for the important OTUs can be found in the online supplement [Table 15].



B	Discriminative OTUs
Prevotella oralis (OTU5)	
Streptococcus (OTU6)	
Prevotella oris (OTU8)	
Porphyromonas (OTU11)	
Flavobacterium succinican (OTU13)	
Haemophilus influenzae (OTU14)	
Bacteroidales (OTU29)	
Elizabethkingia meningoseptica (OTU74)	
Dialister (OTU79)	
Flavobacterium gelidilacus (OTU94)	

Figure 18: Bacterial Microbiome Overview. Phylum relative percent abundances in control and GOLD 4 lung tissue (A). Most important species for discriminating control and GOLD 4 microbiomes, using Random Forest analysis with Boruta feature selection (B). Heatmap of the important bacterial species (C). A) The distribution of phyla was significantly different between control and GOLD 4 ($P<0.05$), and this was driven by Proteobacteria ($P < 0.05$). B) The average (± SD) 10-fold cross validated error rate was 17 ± 2% with a per class error rate of 34 ± 4% for the controls and 6 ± 3% for the GOLD 4 group. C) The samples were clustered by similarity and the three color coded bars at the top represent control or COPD GOLD 4 group, patient, and lung height, respectively.

In order to assess contamination, negative water controls were assessed for the important OTUs identified by Boruta feature selection with random forest analysis for both protocol 1 and 2 [Figure 20 & 21]. Except for *Streptococcus* in protocol 1 all OTUs that were identified as discriminative for control and GOLD 4 lung tissue contained very low square root relative abundances or they were not identified at all in the negative control samples [Figure 20 & 21].

Table 10 summarizes the results obtained by comparing the microbiome data at the phyla level to the host response measured in terms of the Vv of lung tissue occupied by infiltrating inflammatory immune cells. These comparisons show that Shannon diversity is negatively correlated to CD4+ lymphocyte infiltration and also shows Shannon diversity was positively correlated to lung surface area. OTU richness was also negatively correlated with CD4+ lymphocyte infiltration. Further analysis shows that neutrophil infiltration was negatively associated with the presence of Proteobacteria, Comamonadaceae, *Pseudomonas*, and Betaproteobacteria OTUs [Table 10 & 11]. In addition, it also shows that eosinophil infiltration and elastin content were positively associated with Actinobacteria OTUs and B cell infiltration with Propionibacterium, Micrococcaceae and Atopobium OTUs.

Table 10: Summary of Phyla and Diversity Correlations

Significant Result	Coefficient	P-Value	FDR
Shannon Diversity vs. Alveolar CD4 T-cells	-6.99	0.0042	0.05
Shannon Diversity vs. Surface area	0.12	0.014	0.09
OTU Richness vs. CD4 T-cells	-183	0.006	0.06
Proteobacteria vs. Neutrophils	-1.63	0.011	0.09
Actinobacteria vs. Eosinophils	8.63	2.2x10-5	0.0005
Actinobacteria vs. Alveolar B-cells	2.11	0.0025	0.037
Actinobacteria vs. Elastin	0.13	0.0054	0.067

Table 11: Summary of Significant Genus and Species Results

Significant Result	Direction	P-Value	FDR
Comamonadaceae (OTU2) and Neutrophils	Negative	0.005	0.06
Comamonadaceae (OTU49) and Neutrophils	Negative	4.5x10-4	0.019
Pseudomonas (OTU42) and Neutrophils	Negative	0.0042	0.06
Betaproteobacteria (OTU83) and Neutrophils	Negative	0.0058	0.06
Propionibacterium acnes (OTU22) and B-cells	Positive	0.015	0.086
Micrococcaceae (OTU41) and B-cells	Positive	0.026	0.087
Atopobium (OTU98) and B-cells	Positive	0.023	0.057

The data in table 12 summarizes the relationships between predictive OTUs selected by the Random Forest analysis and the results obtained by quantitative histology and microCT. These data show that the Vv of neutrophil infiltration was positively correlated with Dialister (FDR = 0.0001), Bacteroidales (FDR=0.03), *Streptococcus spp.* (FDR =0.06), and *H. influenzae* (FDR=0.06). The number of terminal bronchioles/mL was positively correlated with both *H. influenzae* and *Dialister spp.* (FDR < 0.05). *E. meningoseptica* was positively correlated with Vv of elastin, CD4+ T-cells, and Lm (FDR <0.1) and negatively correlated with total alveolar collagen (FDR = 0.09). *Flavobacterium succinicans* was negatively correlated with CD4 + T-cells while *Flavobactierum gelidilacus* was positively correlated with alveolar surface area (FDR <0.06).

The changes in microbiome composition were associated with a number of host gene expression differences. We identified 859 genes whose expression was associated with the presence of bacteria from the Firmicutes phylum at an FDR cutoff < 0.1 [Table 16]. DAVID analysis indicated that the downregulated genes were mostly involved with Zinc Finger domain regions (FDR < 2.5E-10) while the upregulated genes were involved with pathways with disulfide bonding, signal peptides, membrane, and defense response (FDR < 2E-3). This finding does not change if an FDR cutoff of 0.05 is used instead of 0.10 (data not shown). Additionally, Proteobacteria were associated with 235 genes below an FDR of 0.1 [Table 16]. No pathways were identified from the downregulated genes but the upregulated genes were involved with pathways for splicing, cilium, cell projection, and cell-cell junctions (FDR < 0.1). When the most important predictive bacterial OTUs were analyzed for a correlation with human host gene expression only *H. influenzae* was associated with a single gene below an FDR cutoff of 0.1 (C21orf51, FDR=0.05). The GSEA analysis comparing host gene expression versus the microbiome from protocol 2 to those reported here from protocol 1 shows that the same genes were up and down regulated in relation to Shannon diversity and OTU richness [Table 17, Figure 19, 22-23]. In addition, the analysis based on DAVID of protocol 2 showed that Shannon diversity was positively associated with genes in the dynein, coiled coil, cilium, and microtubule motor activity pathways [FDR < 0.0004] required to clear the mucosal surface. Whereas, negative association between Shannon diversity and gene expression involving genes in the immunoglobulin, glycoprotein, and Fc gamma receptor III pathways that are related to the immune response. The genes used for this analysis can be found in the data appendix.

Table 12: Summary of the Results Obtained of the 10 Discriminative OTUs versus both Structural and Cellular Lung Components.

Comparison	Coefficient	P-Value	FDR
Dialister and Vv of neutrophils	0.32	6.93x10-7	9.01x10-5
E. meningoseptica and Vv of Elastin	0.23	1.58x10-4	0.01
H. influenzae and Number of Terminal Bronchioles	0.01	7.0x10-4	0.02
Flavobacterium gelidilacus and Surface Area	3.0x10-4	6.32x10-4	0.02
Bacteroidales and Vv of neutrophils	0.62	1.24x10-3	0.03
Dialister and Number of Terminal Bronchioles	2.7x10-3	2.49x10-3	0.05
Streptococcus and Vv of neutrophils	1.35	3.65x10-3	0.06
Flavobacterium succinicans and Vv of CD4 T-cells	-2.58	3.96x10-3	0.06
H. influenzae and Vv of neutrophils	0.48	4.35x10-3	0.06
E. meningoseptica and Lm	0.02	4.40x10-3	0.06
E. meningoseptica and Vv of CD4 T-cells	1.00	5.03x10-3	0.06
E. meningoseptica and Vv of Total Collagen	-0.08	8.00x10-3	0.09

4.4 Discussion

The present results confirm earlier reports showing that adult human lungs contain a sparse, yet relatively complex microbiome that maintains density but becomes less diverse in the lungs of patients with COPD (102,103,109). They also extend these observations by showing that both a

touchdown PCR (protocol 1) used to amplify the V3-V5 region and nested PCR (protocol 2) used to amplify the V1-V3 region of the same bacterial 16S rRNA gene showed a decline in OTU richness is association with emphysematous destruction of the lung surface [Figure 17A and 19A]. Both methods also showed differences in the microbial community composition between control lung tissue and tissue from patients with GOLD 4 COPD [Figure 17B & 19B]. In addition they confirm and extend earlier reports (107) by showing [Figure 18] that both the Proteobacteria and to a lesser extent the Actinobacteria expand in COPD as compared to controls whereas the Firmicute and Bacteroidetes phyla contract as the alveolar surface is being destroyed by emphysema in lungs affected by COPD. Most importantly they show that these changes produce a measureable host response in lung tissue.

A recent study by Salter, et al (160), has highlighted the fact that sample contamination is an important source of error in the analysis of sparse yet relatively diverse microbiomes, such as the lung. Therefore it is a concern that some of the OTUs identified as important by the Random Forest analysis in this study (noticeably Flavobacterium and Streptococcus) also do not align to genera identified as potential contaminants (160). Even though the negative controls included with our samples showed these same OTUs were either absent or greatly reduced in our negative control samples [Figure 20 & 21], we cannot conclusively rule out contamination as playing a role in some of the bacteria identified (e.g., Flavobacterium, *E. meningoseptica*, and Dialister). Therefore, these findings need to be interpreted with caution until more precise methods of ruling out contaminating organisms are developed.

A Random Forest analysis showed [Figure 18B] that the OTUs best able to distinguish between lung tissue from control subjects and patients with severe GOLD 4 COPD had both positive and negative effects. For example, the observation that *H.influenzae* is virtually absent in very severe GOLD 4 COPD and increases in association with the numbers of terminal bronchioles observed in the milder forms of COPD could suggest a protective phenomenon. This type of effect has been previously demonstrated in mice: where simultaneous inoculation of *H.influenzae* and *S. pneumonia* onto the upper respiratory mucosal surface showed that *H.infleunzae* out competes *S.pneumonia* for the mucosal surface by inducing a host response that brings in neutrophils to destroy the *S.pneumonia* (180). These observations suggest the hypothesis that *H.influenzae* is capable of causing infection and producing acute exacerbations in the early stages of COPD (142). Moreover it is also consistent with the hypothesis that the decline in terminal bronchioles and increase in emphysematous destruction associated with progression of COPD destroys the habitat that favored the emergence of *H.influenzae* and allows a different set of microbes to emerge, colonize, and infect lung tissue in late stage COPD. Additionally, the tissue vacated by *H.influenzae* might provide a niche for certain exotic bacterium such as *E.meningoseptica* that correlate with inflammatory immune cell infiltration and the tissue remodeling that correlate with progression of COPD in this study. However, additional studies that take into account all of the recently reported corrections for contamination will need to be performed to get the best description of the host response to the microbiome in COPD.

The relative expansion of Proteobacteria, and to a lesser extent Actinobacteria, that occurred in relation to the contraction of the Firmicutes and Bacteroidetes phyla, in this study, is consistent

with a competition for space on the reduced alveolar surface created by emphysematous destruction. For example, the expanded Proteobacteria phylum [Figure 18A] contributed all five of the individual OTUs associated with neutrophil infiltration and 1/4 of the OTUs associated with B cell infiltration [Table 10 & 11]. The smaller expansion of the Actinobacteria phylum contributed 3/4 OTUs associated with B cell infiltration as well as a very strong association with eosinophil infiltration. In contrast, the Firmicute phylum did not contain any OTUs associated with specific responses and the Bacteroidetes phylum only contained *E. meningoseptica* that helped separate the control from COPD GOLD 4 cases. Collectively these data suggest that OTUs located within the phyla that expand as the alveolar surface is destroyed stimulate the host reponse to a greater degree than OTUs in the phyla that contract. Moreover, they suggest the hypothesis that the organisms that compete successfully for the contracting bronchiolar and alveolar surface are recognized by the host immune surveillance system that normally doesn't respond to the bacterial microbiome of the lung.

The gene expression profiling data provide additional evidence in support of a robust host response to changes in the composition of the bacterial microbiome, e.g., by showing that 859 and 235 genes whose expression was either up or down regulated in association with the presence of bacteria from the Firmicutes or Proteobacteria phylum, respectively, at an FDR cutoff < 0.1 [Table 16]. Moreover, the GSEA analysis showed that many of the bacteria associated with changes in host genes were directionally the same using both protocol 1 and protocol 2

Approximately 20 years ago Fredricks and Relman (181) upgraded Koch's postulates for situations where identification of microorganisms is based on sequencing technology. These revised criteria include 1) the nucleic acid sequence of the putative pathogen must be preferentially found in organs or anatomic sites within organs known to be diseased. 2) Fewer or no copies of that sequence should be found in non diseased regions of affected organs. 3) Resolution of the disease should be associated with a decrease in copy number and relapse with increased copy number of the putative pathogen. 4) A causal relationship is more likely if sequence detection predates disease and increases in copy number occur in association with disease progression. 5) The nature of the microorganism inferred from the available sequence data is consistent with the known biological characteristics of that group of organisms thought to be responsible. 6) That *in situ* hybridization techniques be used to demonstrate the relationship between organism and disease at the cellular level. 7) All of the sequence-based forms of evidence for microbial causation should be reproducible.

Although the present results do not satisfy all of these criteria they provide preliminary data showing that OTUs within the expanding Proteobacteria and Actinobacteria phyla account for all the associations observed between individual OTUs and infiltrating inflammatory immune cells. Based on these findings we postulate that the persistent low level inflammatory immune response that has been associated with the progression of COPD (53) is primarily driven by OTUs from within the phyla that expand on a diminishing bronchiolar and alveolar surface with progression of COPD (57). Further, we suggest that the milieu created by these changes allows particular OTUs from within these expanding phyla to punctuate this progressive decline with acute exacerbations of COPD.

An important limitation of this study is that a relatively large number of samples needed to be studied from a small number of individuals, in order to observe the progression of disease within individuals on the same genetic background. The heterogeneity of the disease within individuals and the observation that terminal bronchioles are destroyed prior to the onset of emphysematous destruction makes it possible to assess the response at different levels of tissue destruction (57), but future studies of larger numbers of cases that include better methods of assessing the host response to specific microbial antigens are needed to confirm the present results. Despite this obvious shortcoming the experimental approach described here provides preliminary evidence in support of the hypothesis that there is a host response to the microbiome in COPD and that it is primarily directed at OTUs within the expanding Proteobacteria and Actinobacteria phyla that have successfully competed for space on a reduced alveolar surface. Further, even though none of the patients receiving a transplant had an exacerbation at the time of their transplant, we postulate that the milieu present within the lung microbiome might encourage the emergence of strains from within the expanding Proteobacteria phylum that is known to contribute many of the organisms that produce acute exacerbations of COPD (182,183).

4.4 Online Data Supplement

4.4.1 Consent

Consent was obtained directly from patients treated for very severe COPD by lung transplantation at the Hospital of the University of Pennsylvania and from the next of kin for

unused donor lungs obtained through the Gift of Life Program in Philadelphia. Consent for the use of specimens removed as treatment for lung cancer at St. Paul's Hospital in Vancouver was obtained directly from the patients before the surgery was performed.

4.4.2 Sample Preparation

4.4.2.1 Lung Specimen Preparation and Sampling

The mainstem bronchus was cannulated and the lung was fully inflated with air to a transpulmonary pressure of 30 cm H_2O, then deflated to a transpulmonary pressure of 10 cm H_2O, and held in that position while frozen solid by liquid nitrogen vapor. The specimen was kept frozen while a multi detector computed tomography (MDCT) scan was performed and then cut into approximately 10-14 contiguous 2 cm thick transverse slices between the lung apex and base. The position of each sample was recorded in the MDCT scan of the intact specimen by comparing the photographs of the slices to the corresponding slices of the MDCT scan. This allowed representative samples to be obtained by comparing the CT densities of the samples within each cluster to the distribution of the CT densities of all the samples in the entire specimen.

A cluster of 4 cores of tissue approximately 1.2 cm in diameter and 2cm in length located in close proximity to each other were removed from each of these lung slices. One sample from each cluster was assigned to microCT examination (57), the 2nd for gene expression profiling (19), the 3rd to quantitative histological analysis, and the 4th to bacterial 16S ribosomal DNA sequencing and microbiome analysis.

Additionally, lungs for both the controls and COPD GOLD 4 were all processed at the same institution by the exact same people. Likewise the slicing and coring of the lung were performed by the exact same people and in the same room for every lung. After samples were cored only a piece of tissue from the middle of the sample was utilized for DNA extraction and the other side was used for tissue printing culture. The procurement of tissue for DNA extraction and tissue printing occurred in a biosafety cabinet level 2.

4.4.2.2 MicroCT

The samples processed for microCT were kept at -80ºC while fixed in a solution of pure acetone and 1% gluteraldehyde (freezing point -92ºC) overnight. The fixed specimens were then warmed to room temperature and processed for microCT examination as previously described (57). The analysis of the electronic record of the microCT scans provided measurements of the number of terminal bronchioles/ml tissue in each core and the mean linear intercept (Lm) measured at 20 different levels within each tissue core.

4.4.2.3 Quantitative Histology

Portions of the 53 cores of tissue assigned to histology were vacuum embedded in a solution (50%v/v Tissue-Tek O.C.T. (Sakura Finetek USA Inc, Torrance, CA, USA) in PBS with 10% sucrose) at 1°C and immediately refrozen on dry ice. Cryosections cut from these blocks stained with hematoxylin and eosin (H&E) were examined microscopically to establish that 42 of the 61

frozen sections contained bronchioles suitable for histological analysis. Bronchiolar wall thickness was measured from histological sections as previously described (53). Additional sets of serial histological sections were immunostained with appropriate antibodies to identify type I collagen (collagen I), type III collagen (collagen III), elastin, macrophages, CD4+ and CD8+ T cells, B cells, natural killer (NK) cells, and polymorphonuclear neutrophils (PMNs). Eosinophils were identified by Hansel's stain and picrosirius red staining was used to identify total collagen. Digital images of the histological sections were stored and subsequently analyzed using Image-Pro Plus software (Media Cybernetics, Bethesda, MD, USA) (53).

4.4.2.4 Gene Expression Profiling

High molecular weight (HMW; mRNA-containing fraction) RNA was isolated from each tissue core using the miRNeasy Mini Kit (Qiagen) and assigned to gene expression profiling. The RNA integrity was assessed using an Agilent 2100 Bioanalyzer and RNA purity was assessed using a NanoDrop spectrophotometer. One gram of RNA was processed and hybridized onto the Human Exon 1.0 ST array (Affymetrix Inc.) according to the manufacturer's protocol as previously described (184). Expression Console Version 1.1 (Affymetrix Inc.) was used to generate transcript-level gene expression estimates for the "core" exon probe sets via the robust multichip average (RMA) algorithm. Gene symbols of transcript IDs were retrieved using DAVID (http://david.abcc.ncifcrf.gov/) (178).

4.4.2.3 Bacterial Microbiome Analysis

Total bacteria qPCR (109) and pyrotag sequencing (185,186) of the 16S bacterial ribosomal DNA were performed on DNA from 69 lung tissue cores, on 8 negative controls where sterile water was used in place of the sample DNA and on five technical replicates of randomly chosen control lung samples. A total of 29 control samples and 40 samples from subjects with very severe COPD were analyzed.

4.4.2.4 Touchdown PCR Approach with V3-V5 Primers

The protocol is reported in PLoS One (175). Briefly, pyrosequencing flowgrams were trimmed to 450 flows and denoised using shhh.flows function in mothur. Denoised reads were filtered further by removing sequences containing homopolymers longer than 8 bp and extraction of sequences at least 200 bp in length. Sequences were aligned using the SILVA database (http://www.arb-silva.de/) and then organized so that sequences would all be approximately the same size. Chimeras were removed using UCHIME and any reads that did not align to bacteria were also removed. The remaining reads were clustered into OTUs based on a ~97% similarity threshold utilizing the dist.seqs command in mothur. This was the processing used to prepare data for all figures in the earlier sections of this chapter.

4.4.2.5 Nested PCR Approach with V1-V3 Primers

Pyrosequencing flowgrams were first trimmed to 450 flows and denoised with PyroNoise (187). Denoised reads were filtered further by removing sequences containing homopolymers > 6 bp, followed by extraction of variable regions V1 and V2 with *V-Xtractor* (188). Chimeras were removed using UCHIME (189), and the remaining reads clustered into operational taxonomic units (OTUs) delineated by a ~97% similarity threshold using a Levenshtein-distance-based algorithm [*crunchclust* (http://code.google.com/p/crunchclust/)] (190). We obtained a consensus taxonomy for each OTU (using a 50%-majority voting scheme) after taxonomically classifying reads using the Bayesian method with a bootstrap support of ≥ 80%) (191). To increase classification depth, sequence classification was performed against the full Greengenes database (192) trimmed, via *V-Xtractor*, to regions V1-V2 (193). Additionally, the SILVA database was used instead of the Greengenes database for classification to examine the effect of database usage on the final results. Except for the *V-Xtractor* trimming, all steps were performed using *mothur* v. 1.27 (194) and software implementations therein.

4.4.3 Data Analysis

4.4.3.1 MicroCT

MicroCT was analyzed using a previously described modification of a multi-level cascade sampling design where the reference volume for the entire lung was computed from the HRCT scans of the intact specimen, and the sub volumes present in each tissue core were measured from the electronic record of their microCT scan. This allowed terminal bronchioles to be

identified anatomically within each tissue core, their number /ml of sample counted and their lumen diameter measured (57). The product of the mean number of terminal bronchioles/ml lung and total lung volume provided the total number of terminal bronchioles /lung. In addition the comparison of the number of terminal bronchioles / ml to the mean linear intercept (Lm) measured at 20 equidistant intervals from top to bottom of each tissue core allowed the number of terminal bronchioles in each tissue core to be compared to the emphysematous destruction present in that core of tissue. Also the Lm obtained by microCT was used to compute the alveolar surface area (SA) of each tissue core using the formula: SA = 4 x V/Lm where V = the volume of the tissue core.

4.4.3.2 Microarray Analysis

Two linear mixed-effects models were used to identify gene expression profiles associated with bacteria within the microbiome, with structural lung components, or volume fractions of cellular infiltrates with the latter three denoted as lung components (LC).

Equation (1) representing the first model is the following

(1) $Gene_{ij} = {}_0 + {}_{Slice} * Slice_{ij} + {}_j + {}_{ij}$

Where $Gene_{ij}$ is the $\log_2$ expression value for sample *i* in patient *j* for a single gene. Slice is a fixed effect controlling for the position within the lung from which the sample core was obtained. The random term ${}_{ij}$ represents the random error which was assumed to be normally distributed, ${}_j$ represents the random effect for patient, and ${}_0$ represents the intercept.

Equation (2) representing the second model is:

(2) $Gene_{ij} = {}_0 + {}_{Slice}*Slice_{ij} + {}_{LC}*LC_{ij} + {}_j + {}_{ij}$
$i=1,2,\ldots,8$; $j=1,2,\ldots,8{}_{ij} \sim N(0, {}_2)$ ${}_j \sim N(0, {}_{2j})$

The model in equation (2) contains an additional fixed effect term for LC (e.g. macrophages, Ralstonia, etc.). A gene's expression profile was considered associated with a particular LC if model (2) fit better than model (1) as determined by a significant p-value from a likelihood ratio test between the two models after applying a false discovery rate (FDR) correction. This approach has some limitations, for the gene expression, in that not every value of LC to be considered against every other value of LC although it allows for all possible values of a specific LC to be considered by slice within the same individual. However, for the quantitative histology no such limitation occurred with this model and it considered all combinations of correlations between the Vv and important bacterial OTUs. All statistical analyses were conducted using R statistical software v 2.9.2 and the nlme package in Bioconductor v2.4 (195).
Functional enrichment analysis was performed using DAVID 2008 (178). For DAVID, functional enrichment was examined among Gene Ontology (GO) categories, and KEGG and BIOCARTA pathways. All genes in the species *Homo sapiens* were used as a reference set.

4.4.3.3 Quantitative Histology

The volume fraction (Vv) of each cell and tissue type present within the bronchiolar and alveolar tissue as determined on the stored digital images of the stained histological sections were

inserted into a multi-level cascade sampling design to compute the accumulated volume of infiltrating inflammatory immune cells, as well as the volume of each tissue type present at different levels of emphysematous destruction.

4.4.3.4 Bacterial Microbiome Analysis

For both protocols the total number of reads for each community (lung core) was normalized, using random sub-sampling, to the smallest number of reads among the samples after denoising [Table 13], to control for differences in sequencing depth before alpha diversity (observed OTU richness; *S*) and community similarity analyses. Community similarity was visualized with principal coordinates analyses (PCoA) ordination of pair-wise Bray-Curtis dissimilarities computed from square-root transformed OTU relative abundances. The effects of disease status, that is control versus COPD GOLD 4; patient effects; and the position of core sample in the lung, on bacterial community composition differences were statistically examined with a permutational multivariate analysis of variance (PERMANOVA; (148)), which enabled us to quantify the relative proportion of variability explained by each source of variation in the model (149). Ordinations and PERMANOVA were performed with vegan in R for protocol 1 and with the PIMER-E software for protocol 2.

For both protocols further OTU-level analyses, the OTU abundance table was filtered to exclude OTUs with a cumulative summed abundance of $\leq$5 reads. All downstream analysis was performed on the square root relative abundance for each OTU. To identify OTUs that discriminate between control and GOLD 4 communities, we used the Random Forests (RF) algorithm, an ensemble-based supervised classification method that generates multiple weak

classifier decision trees (151). The classification error rate was measured by out-of-bag (OOB) estimation for each group. An importance measure was calculated for each feature (OTU) based on the loss of accuracy in classification when the OTU was removed from analysis. The importance measures was then determined using the Boruta package, a feature selection algorithm built around the Random Forest algorithm (152). RF and Boruta analyses were performed with the Genboree Microbiome Toolset (196). Since error rate was similar regardless of whether RF included negative controls or not (data not shown), we excluded negative controls for selection of discriminative OTUs.

Table 13: List of Reads per Sample for both Protocol 1 and 2

Patient	Core	Protocol 1 Reads per Sample	Protocol 2 Reads per Sample
6965	9	1809	4259
6965	4	1351	5977
6965	6	1685	4458
6965	5	1266	5285
6965	8	1517	3933
6965	10	2006	4899
6965	7	1111	4573
6965	3	918	4336
6969	11	1834	5836
6969	9	1389	3005
6969	4	1814	5684
6969	7	2185	7116
6969	5	1525	5370
6969	2	1358	6822
6969	8	1697	6730
6971	3	822	5566
6971	2	918	6210
6971	11	1785	4520
6971	7	1378	7386
6971	5	1335	6964
6971	8	1305	7585
6971	6	1245	6494
6971	4	954	6022
6978	8	1518	5124
6978	10	1333	6147
6978	7	2020	4756
6978	6	851	6505
6978	9	1421	5415
6978	3	1637	5834
6978	4	1893	7753
6978	5	654	3967
6982	11	1821	5337
6982	2	958	6097
6982	7	4031	2113
6982	3	1111	5527
6982	9	2891	3210
6982	4	1348	3551

Patient	Core	Protocol 1 Reads per Sample	Protocol 2 Reads per Sample
6982	6	1618	2980
6983	3	2428	8399
6983	6	2399	3743
6983	9	1716	4383
6983	4	1944	6414
6983	2	802	9334
6983	7	2174	4277
6983	10	1875	5072
6983	5	2307	4006
6989	3	883	6783
6989	10	2174	5443
6989	6	777	5202
6989	11	2896	6288
6989	9	799	6408
7010	8	1787	6127
7010	7	1623	8816
7010	10	1923	8312
7010	4	1700	7532
7010	3	1097	8148
7010	5	1751	6721
7010	6	2956	10751
7010	2	1363	6957
7014	5	1244	7293
7014	6	1015	7825
7014	4	866	8302
7014	7	1181	9884
7014	2	1171	7286
7014	3	1393	8054
7014	8	1054	7344
7014	9	1310	5594

Table 14: Breakdown of Cases and Cores Used in the Different Analysis of the Study

	Total Cases		Total Cores	
Analysis	**Controls**	**GOLD 4**	**Controls**	**GOLD 4**
microCT	4	3	29	24
Gene Expression	2	3	16	24
Quantitative Histology	4	3	29	24
Bacterial Microbiome	4	5	29	40

Table 15: Summary of Random Forest with Boruta Feature Selection for Important OTUs for the Discrimination between Control and GOLD 4 Based on Database Used.

Greengenes Protocol 2	SILVA Protocol 2	SILVA Protocol 1
Ralstonia	Prevotella oris	Prevotella oralis
Streptococcus pseudopneumonia	Prevotella melaninogenica	Streptococcus
Prevotella oris	Streptococcus	Prevotella oris
Streptococcus	Streptococcus pseudopneumonia	Porphyromonas
Flavobacterium	Ralstonia	Flavobacterium succinicans
Streptococcus anginosum	Gemella	Haemophilus influenzae
Porphyromonas	Prevotella histicola	Bacteroidales
Prevotella	Prevotella	Elizabethkingia Meningoseptica
Streptococcus constellatus	Fusobacterium	Dialister
Prevotella melaninogenica	Streptococcus constellatus	Flavobacterium gelidilacus
Gemella	Porphyromonas	
Veillonella parvus	Flavobacterium	
Propionibacterium acnes		
Prevotella histicola		
Fusobacterium		
Sphingomonas asaccharolytica		
Sphingobium yanoikuyae		
Streptococcus		

Protocol 2 had a significantly higher richness and Shannon Diversity (P<0.0001) and a lower evenness (P<0.0001) versus protocol 1. There was poor direct correlation between the two methods with evenness having the best correlation (R = 0.58) and richness having the worst (R = 0.28). Both control and GOLD 4 samples had sparse bacterial communities, with densities, as measured by qPCR, that were not different (P>0.05, data not shown) and were in the range of those previously reported in lung tissue (109). With respect to detecting bacteria that can commonly cause exacerbations, protocol 1 performed much better than protocol 2. Protocol 2 detected a single read of *Streptococccus pneumonaie*, 7 reads for *Haemophilus influenzae*, and 0 reads for *Moraxella catarrhalis*. In contrast, protocol 1 detected a single read of *Streptococcus pneumonaie*, 1258 reads of *Haemophilus influenzae*, and 0 reads for *Moraxella catarrhalis*.

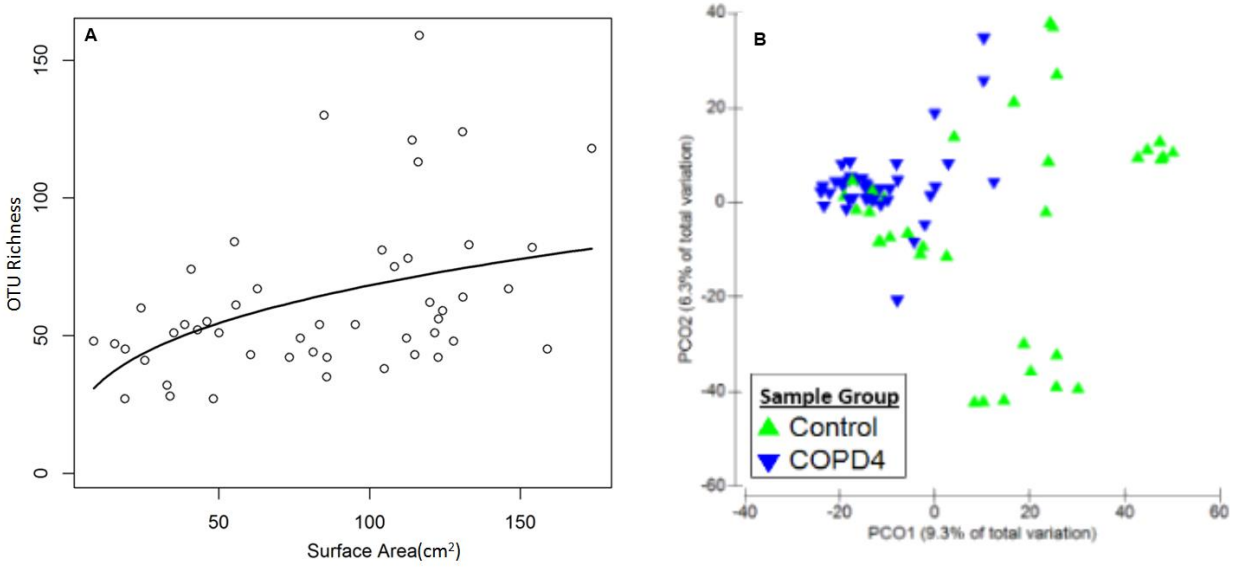


Figure 19: Nested PCR Approach with V1-V3 Primers (Protocol 2)

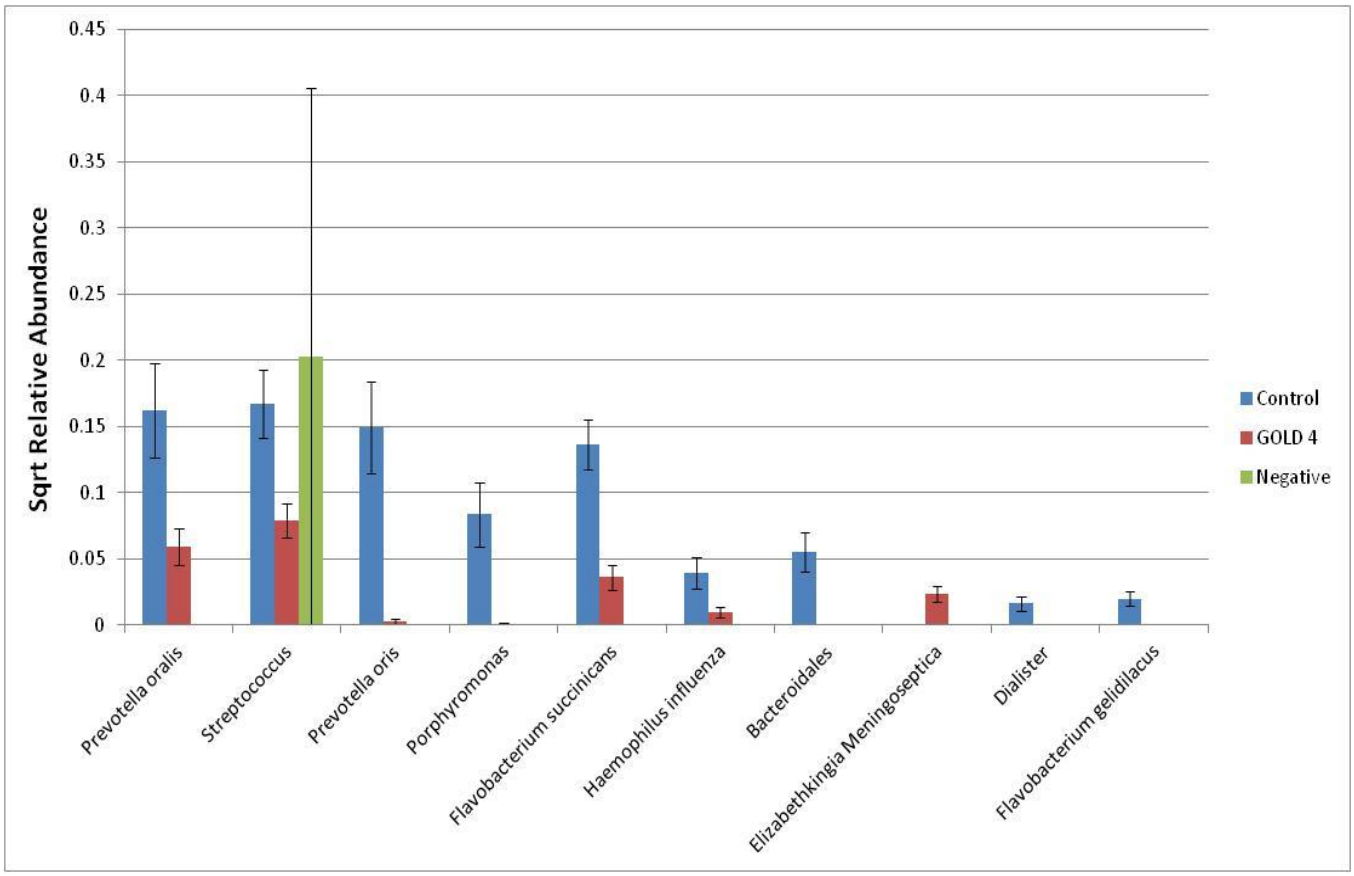


Figure 20: Protocol 1 Overall Breakdown between Control, GOLD 4, and Negative Controls. The top 10 most discriminative OTUs for control (n=29) and GOLD 4 (n=40) versus negative water controls (n=2) run alongside the samples. For all of the top 10 except Streptococcus the OTU was absent from our negative controls. Results reported as mean ± SEM.

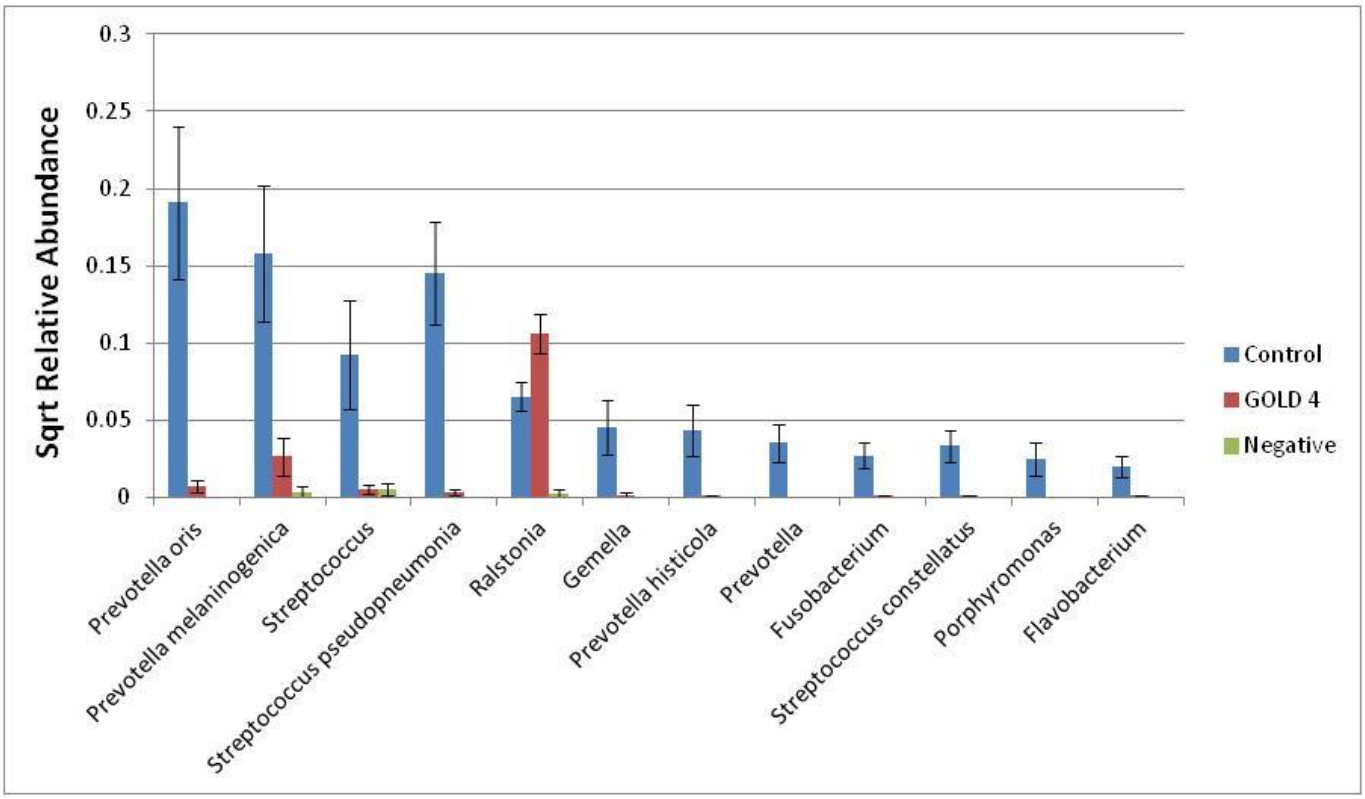


Figure 21: Protocol 2 Overall Breakdown between Control, GOLD 4, and Negative Controls. The top 12 most discriminative OTUs for control (n=29) and GOLD 4 (n=40) versus negative water controls (n=8) run alongside the samples. The top 12 OTUs were either absent or much lower than samples for the negative controls with the exception of Streptococcus being close to equivalent between negative controls and GOLD 4. Results reported as mean ± SEM

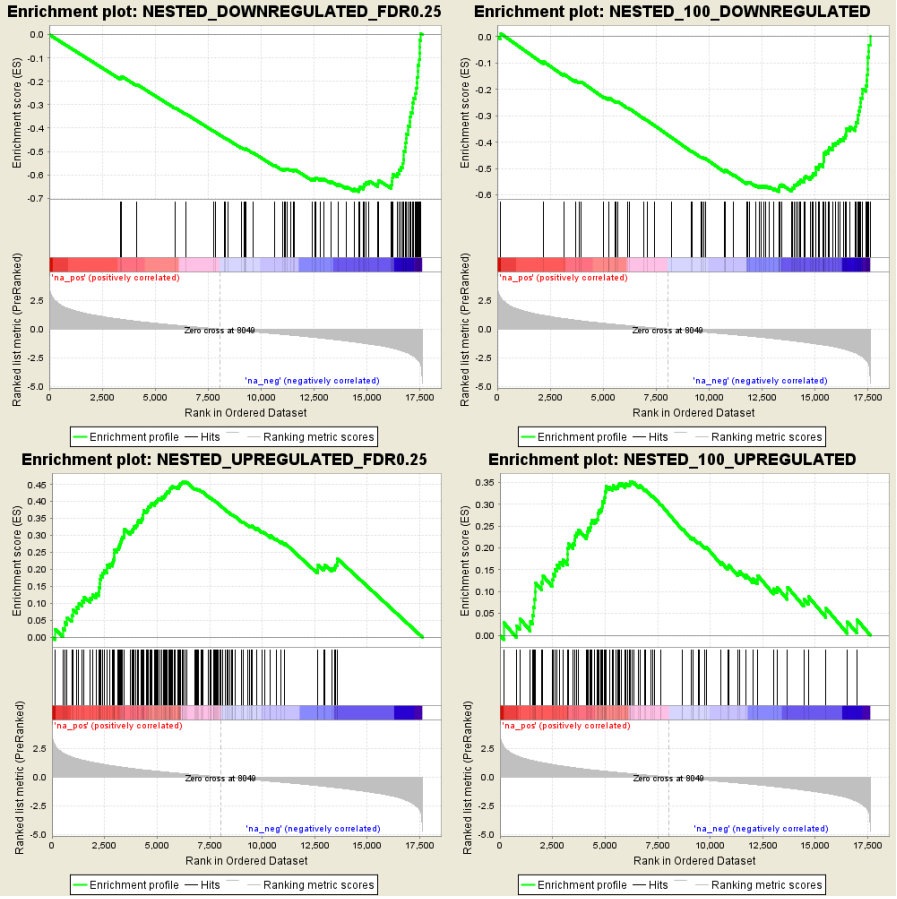


Figure 22: GSEA of Shannon Diversity. Top two panels are downregulated genes and the bottom two panels are for upregulated genes. Genes were taken that were below an FDR of 0.25 or from a random selection of 100 genes from both protocol 1 and 2.

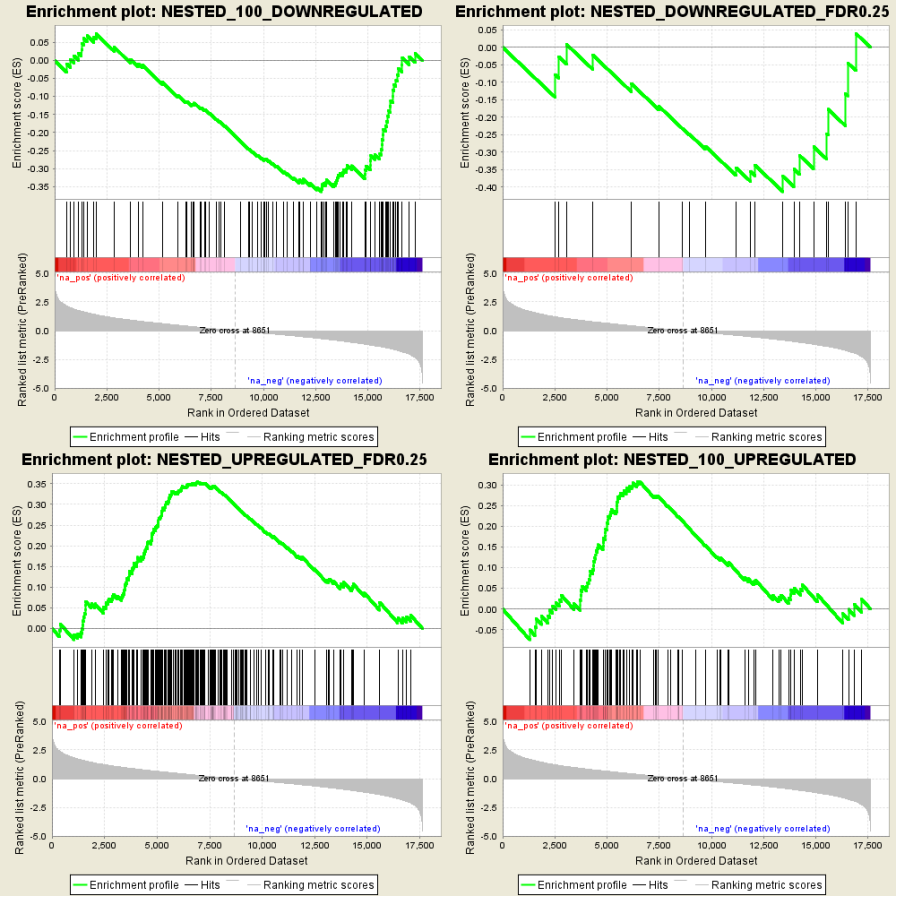


Figure 23: GSEA of OTU Richness. Top two panels are downregulated genes and the bottom two panels are for upregulated genes. Genes were taken that were below an FDR of 0.25 or from a random selection of 100 genes from both protocol 1 and 2.

Table 16: Top 10 Pathways Identified by DAVID from Genes Correlated with either Firmicutes or Proteobacteria

Top 10 Human Pathways Identified by DAVID Correlated with Firmicute Number		**Top 10 Human Pathways Identified by DAVID Correlated with Proteobacteria Number**	
Downregulated	**Upregulated**	**Downregulated**	**Upregulated**
Zinc Finger, C2H2-type	Disulfide Bond	No Pathways Identified	Alternative Splicing
Zince Finger, C2H2-like	Signal		Splice Variant
zinc finger region: C2H2-type 11	Glycoprotein		Cilium Membrane
zinc finger region: C2H2-type 10	Glycosylation site: N-linked (GlcNAc)		Cell-cell Junction
zinc finger region: C2H2-type 4	Signal Peptide		Cell Projection Membrane
zinc finger region: C2H2-type 6	Disulfide Bond		Cilium Part
zinc finger region: C2H2-type 7	Topological domain: Extracellular		Cilium
zinc finger region: C2H2-type 1	Topological domain: Cytoplasmic		Cell Projection
zinc finger region: C2H2-type 3	Membrane		BBSome
zinc finger region: C2H2-type 9	Defense Response		Cell Projection Part

Table 17: Gene Set Enrichment Analysis of Shannon Diversity and OTU Richness between Protocol 1 and Protocol 2.

	Shannon Diversity		OTU Richness	
	FDR	**Enrichment Score**	**FDR**	**Enrichment Score**
Upregulated (FDR < 0.25)	0	0.46	0	0.36
Upregulated Random 100 Genes	0.0028	0.35	0.045	0.31
Downregulated (FDR < 0.25)	0	-0.67	0.066	-0.41
Downregulated Random 100 Genes	0	-0.59	0.016	-0.36

Within this chapter it has been shown that changes in the bacterial microbiome can correlate with both structural and inflammatory components that are important in the progression of COPD. Further, a specific set of 10 OTUs were identified in being able to discriminate between control and COPD GOLD 4 lung tissue samples. In the next chapter I will present data that shows how ddPCR is better than qPCR in detecting a low concentration of 16S rRNA. This will add a valuable perspective when approaching Chapter 6 where a high sensitivity assay was needed to explore how *H.infleunzae*, one of the 10 discriminative OTUs, could potentially drive the severity of COPD.

Chapter 5: Droplet Digital PCR in the Analysis of Bacterial Load in Lung Tissue[3]

5.1 Introduction

Recently, we reported that lung tissue samples of smokers, non-smokers and those with, chronic obstructive pulmonary disease (COPD), and cystic fibrosis (CF) showed increased bacterial population as compared with controls (109). We used qPCR quantitation of 16S rRNA to detect levels of bacterial microbiome in these samples. For absolute quantitation of 16S rRNA, serial dilution of *Escherichia coli* (*E.coli*) DNA was required for generation of a standard curve on every plate. This process can be time consuming and costly, and limits sample throughput. Moreover, one needs to ensure that the standard curve is optimized and contains an effective dynamic range for accurate quantitation of target genes in desired samples (197). Often, results could be misleading as the reaction efficiency of the standard samples may vary from the reaction efficiency of test samples due to differences in sample content and presence of inhibitors (198,199). The requirement for a large number of technical replicates when assessing low abundance genes is another major hurdle associated with this technique, which could be problematic when amount of sample is limited (200). The concentration of 16S rRNA in lung tissue samples is extremely low (1-10 copies/μL), and very close to the lower detection limit of qPCR. Precise and accurate measurement of the low copies of 16S rRNA in lung tissues is

[3] This section has been previously published in PLOS One.
Sze MA, Abbasi M, Hogg JC, Sin DD. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue from control and COPD GOLD 2. *PLOS One*. 2014, 9(10):e110351.doi:10.1371/journal.pone.0110351

essential to differentiate between negative controls, smokers, non-smokers, COPD, and CF samples. For this purpose, a more precise method is required for 16S rRNA quantification.

Droplet digital™ PCR (ddPCR) allows for absolute quantitation of nucleic acids without the requirement for standard curves. The technique is based on partitioning of a single sample into 20,000 much smaller, segregated reaction vessels (known as droplets). A standard PCR reaction can then be employed to amplify the target(s) in each droplet which can be individually counted by the associated target dependant fluorescence signal as positive or negative. The simple readout of droplet partitions as a binary code of ones (positive) and zeroes (negative) represents the "digital" aspect of the technique and because the presence of a target in a given droplet is a random event, the associated data fits a Poisson distribution (201,202). This permits the direct and simple calculation of DNA copy numbers in a sample without the requirement of a standard curve. Since ddPCR is an end point PCR reaction, data are not affected by variations in reaction efficiency and as long as the amplified droplets display increased fluorescence intensity compared to the negative droplets, absolute copy number of target genes can be obtained with a high degree of confidence. Owing to the high precision and accuracy of this technique, the need for technical replicates is reduced (203), and the Poisson distribution provides 95% confidence intervals for measured copies from single wells which provides robust estimates of data dispersion obtained from technical replicates (204). This can significantly increase sample throughput, save time, and effectively allow accurate quantitation of precious samples.

Sample partitioning in ddPCR also improves sensitivity when quantifying low concentration of target genes in a highly concentrated complex background (203,205,206). When quantifying a low amount of 16S bacterial rRNA in DNA extracted from human lung tissue, the 16S primers have a difficult task of browsing through the large number of non-specific sequences contained in the complementary strand. This reduces sensitivity of the assay by introducing noise in target amplification. By using ddPCR to partition sample into 20,000 droplets we are able to increase the signal to background ratio by a factor of 20,000 and the primers and probes are able to locate the target sequence from a far less concentrated background. Using this technique, we aim to increase accuracy and sensitivity in detecting total bacteria within the lung of smokers, non-smokers, and COPD patients.

5.2 Methods

5.2.1 Tissue Samples

Lung tissue was obtained from the tissue registry at St. Paul's Hospital. Ethics approval was specifically obtained for this study from the University of British Columbia - Providence Health Care (UBC-PHC) Research ethics board. Informed consent was obtained, through a written consent form, and approved by the UBC-PHC Research ethics board for patients who underwent lung resection therapy for various pulmonary conditions, such as lung cancer, for collection and use in this study. For this study, we used lung tissue from the tumor-free part of the resected lung segment. Samples were obtained from 16 control (FEV1/FVC >0.7) and 16 patients with moderate COPD GOLD 2 (Global initiative for chronic Obstructive Lung Disease) (FEV1/FVC

< 0.7, and 50% < FEV1 < 80%) were used. Resected lung tissues were inflated with cryomatrix (OCT) and then frozen in liquid nitrogen. From this, 2cm thick contiguous transverse slices were made and tissue samples were taken from one of these slices. Frozen sections were obtained by cutting the tissue sample on a cryotome with some sections assigned for DNA extraction and others used for quantitative histology (144).

5.2.2 Experimental Protocol

DNA from all samples was extracted using a Qiagen DNeasy Extraction kit according to the manufacturer's instructions and the concentration was assessed using Nanodrop. qPCR (Applied Biosystems ViiA7) was performed on these samples using a previously published 16S rRNA assay (109) that utilized a standard curve of a serial dilution of *Escherichia coli* (cycling conditions were 1 cycle at 95ºC for 15 minutes, 40 cycles at 95ºC for 15 seconds and 60ºC for 1 minute, followed by a standard denaturation curve protocol). The assay was a SYBR green qPCR assay and three replicates were used per sample. Data were collected using the ABI ViiA7 RUO software program. The same assay was adapted to ddPCR (Bio-Rad QX200^{TM}) and the experiments were performed using the following protocol: 1 cycle at 95ºC for 5 minutes, 40 cycles at 95ºC for 15 seconds and 60ºC for 1 minute, 1 cycle at 4ºC for 5 minutes, and 1 cycle at 90ºC for 5 minutes all at a ramp rate of 2ºC/second. Bio-Rad's T100 thermal cycler was used for the PCR step. No standard curve was required for the ddPCR and the droplets were quantified using the Bio-Rad Quantisoft software. A total of two replicates were used per sample. A threshold cutoff of 20000 was chosen based on preliminary experiments, which accurately

separated positive from negative droplets. For both protocols, negative controls that comprised of DNase and RNase free water were used and were run alongside the samples.

5.2.3 Quantitative Histology

Sections were stained with Movat pentachrome stain and Hematoxylin and Eosin (H&E) to obtain the mean linear intercept (Lm) which is a marker of emphysematous destruction of airspaces (56,57). The arithmetic mean of the Lm obtained from the Movat pentachrome and H&E stained sections were used as the analytic value of Lm for each tissue sample. Immunohistochemical staining for both the small airway and alveolar volume fraction (Vv) of CD4 T-cells, CD8 T-cells, B-cells, macrophages, and neutrophils (PMN) were obtained by using a grid based point counting method to obtain a positive cell:tissue ratio for each cell type (53).

5.2.4 Data Analysis

Analysis involved testing whether ddPCR and qPCR protocols could differentiate 16S in tissue samples from those of the negative controls. Direct comparison of the total 16S obtained with both methods was made to detect differences between tissue samples and negative controls. The coefficient of variation between ddPCR and qPCR methods was then compared. Finally, the data generated from both techniques were compared with important histological measures of COPD to determine the relationship of 16S findings from ddPCR and qPCR with parameters of COPD. Grouped analysis used Kruskall-wallis ANOVA analysis with Tukey's post hoc testing. Standard t-tests were used in non-grouped analysis. Using a standard test statistic for

significance of the regression line was tested for its difference from 0. A P-value < 0.05 was considered statistically significant and all analysis was performed using Prism v. 5 (GraphPad Software Inc. La Jolla California).

5.3 Results

5.3.1 16S Detection with qPCR or ddPCR

Both qPCR and ddPCR were assessed for their ability to detect 16S and whether the samples were above the negative non-template control samples. Figure 24A shows that the qPCR assay was able to detect the bacterial 16S rRNA gene and that both controls and moderate COPD samples were significantly higher than that of the negative controls ($P < 0.05$). Figure 24B shows that the ddPCR could also detect the bacterial 16S rRNA gene and that both controls and moderate COPD samples were significantly higher than the negative controls ($P < 0.05$). Both methods showed no significant difference in total 16S bacterial load between control and moderate COPD ($P > 0.05$).

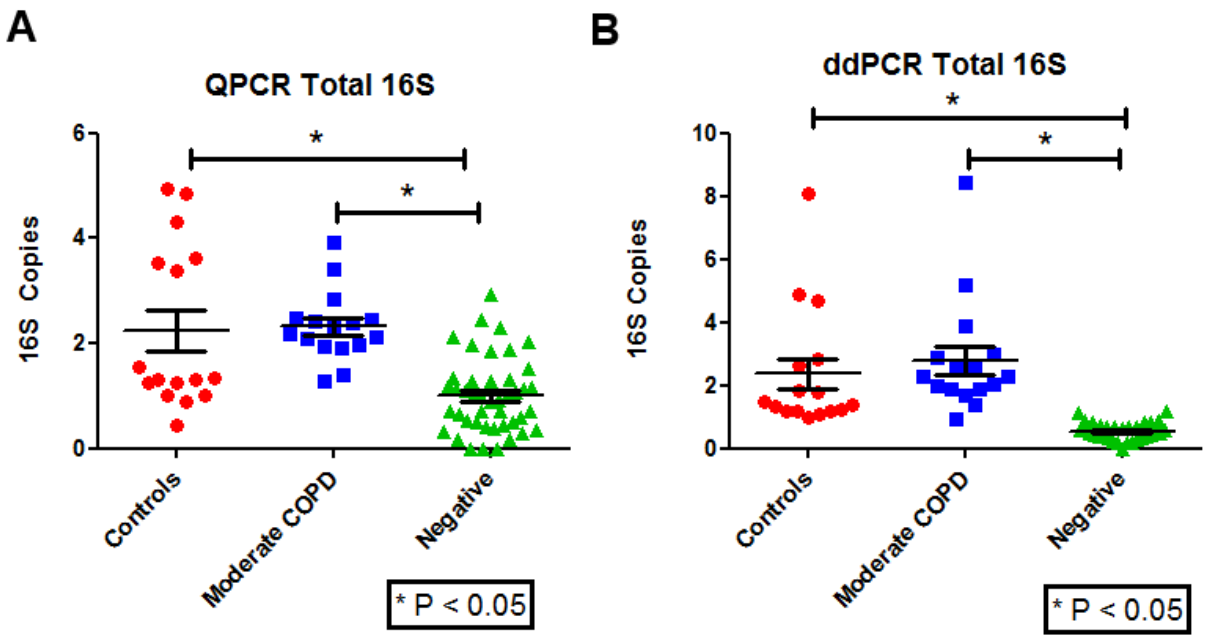


Figure 24: Head to Head Comparison of QPCR and ddPCR 16S Quantification. Bars represent the mean ± SEM and an ANOVA with a Tukey pot hoc test was utilized for statistical analysis.

5.3.2 Comparison of the qPCR to ddPCR 16S rRNA Assay

The ddPCR negative controls had a much smaller standard deviation versus the qPCR negative controls (0.28 versus 0.70). Both the qPCR and ddPCR detected a similar bacterial load for the control and moderate COPD groups. For the moderate COPD group, qPCR values were 2.32 ± 0.67 16S copies (mean ± SD) and ddPCR values were 2.80 ± 1.80 16S/uL and for

the control group the qPCR and ddPCR values were 2.25 ± 1.55 16S copies and 2.36 ± 1.95 16S/uL (mean ± SD) respectively. There was a significant decrease in the negative control 16S bacterial load using the ddPCR technique compared with qPCR ($P < 0.0032$). The ddPCR had a value of 0.55 ± 0.28 16S/uL and the qPCR having a value of 1.00 ± 0.70 16S copies [Figure 25A]. There was a significant positive relationship between the qPCR and ddPCR 16S counts with an R^2 value of 0.27 [Figure 25B]; the line of best fit was $y = 0.33x + 1.44$. Further, the ddPCR coefficients of variation (CV) were significantly lower than those obtained by the qPCR assay (P-value < 0.0001) [Figure 26A]. The average CV for the ddPCR was 0.18 ± 0.14 while for the same samples the CV for the qPCR was 0.62 ± 0.29. Using a Bland-Altman plot to further analyze the CV data and using the ddPCR as the reference against the qPCR the bias was found to be -0.44 ± 0.29. This means that on average for any given sample the qPCR CV will be 0.44 ± 0.29 higher than the ddPCR CV [Figure 26B].

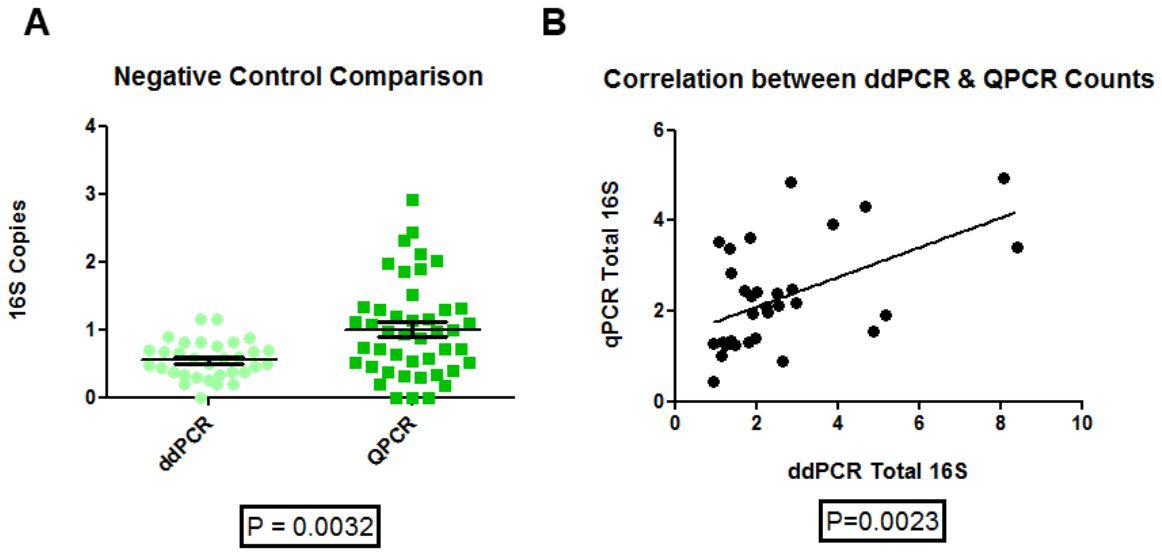


Figure 25: Negative Control and Direct 16S ddPCR versus qPCR Comparisons. The bars represent the mean ± SEM. A standard T-test was utilized for panel A while a standard test statistic was used to test departure from zero in panel B.

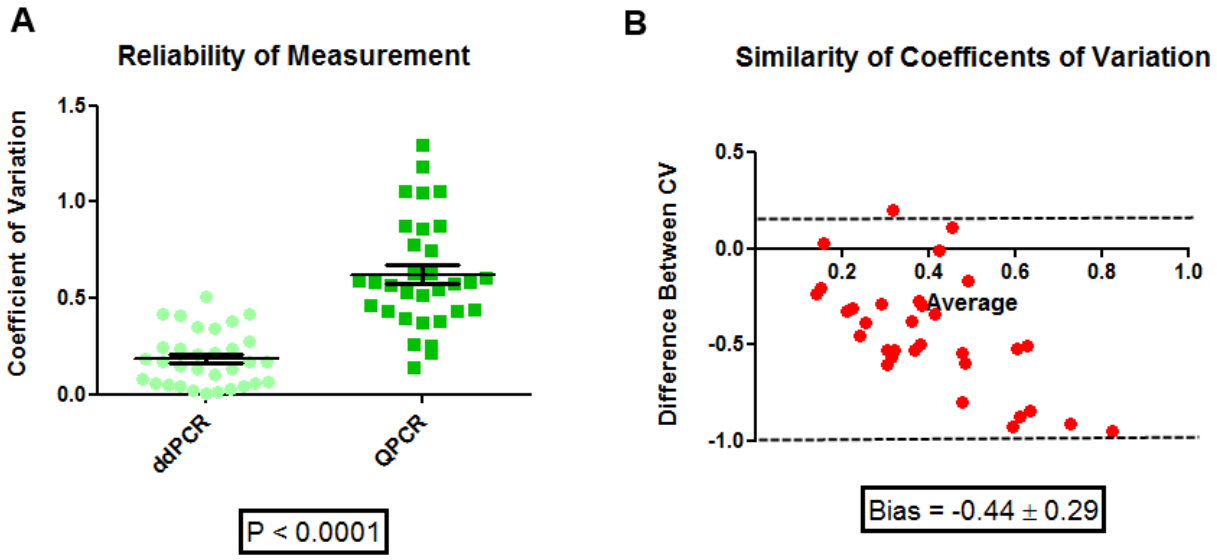


Figure 26: Comparison of the Coefficients of Variation (CV) between qPCR and ddPCR. The bars represent the mean ± SEM. A T-test was used for the analysis of panel A and a Bland-Altman plot was used for panel B.

5.3.3 Comparison of the qPCR to ddPCR 16S rRNA Assay and Correlations to Important Tissue Measurements of COPD

Using quantitative histology [52], we examined the relationship between ddPCR values and qPCR values and parameters of tissue remodeling and lung inflammation, which are salient histologic features of COPD. Both methods generally were similar when there was no significant correlation between the 16S counts and histologic measurements for tissue remodeling or inflammation ($P>0.05$). However, correlations with CD4 that was not previously significant using qPCR became significant when we used ddPCR [Figure 27C & D]. When there were significant correlations ($P<0.05$) ddPCR data was more tightly associated with the histologic measures of tissue remodeling compared with qPCR [Figure 27A & B]. Overall

ddPCR demonstrated a larger slope than qPCR and also tended to have a greater dynamic range [Figure 27]. Figure 30A and 30B show the improved correlation between emphysematous tissue destruction (Lm) and total 16S bacterial counts with ddPCR (P-value < 0.0001, $R^2 = 0.54$) versus qPCR (P-value = 0.015, $R^2 = 0.19$). Similarly, figure 28C and 28D show the improved correlation between infiltration CD4 T-cells into the alveolar tissue and total 16S bacterial counts with ddPCR (P-value = 0.0004, $R^2 = 0.69$) versus qPCR (P- value = 0.242, $R^2 = 0.12$).

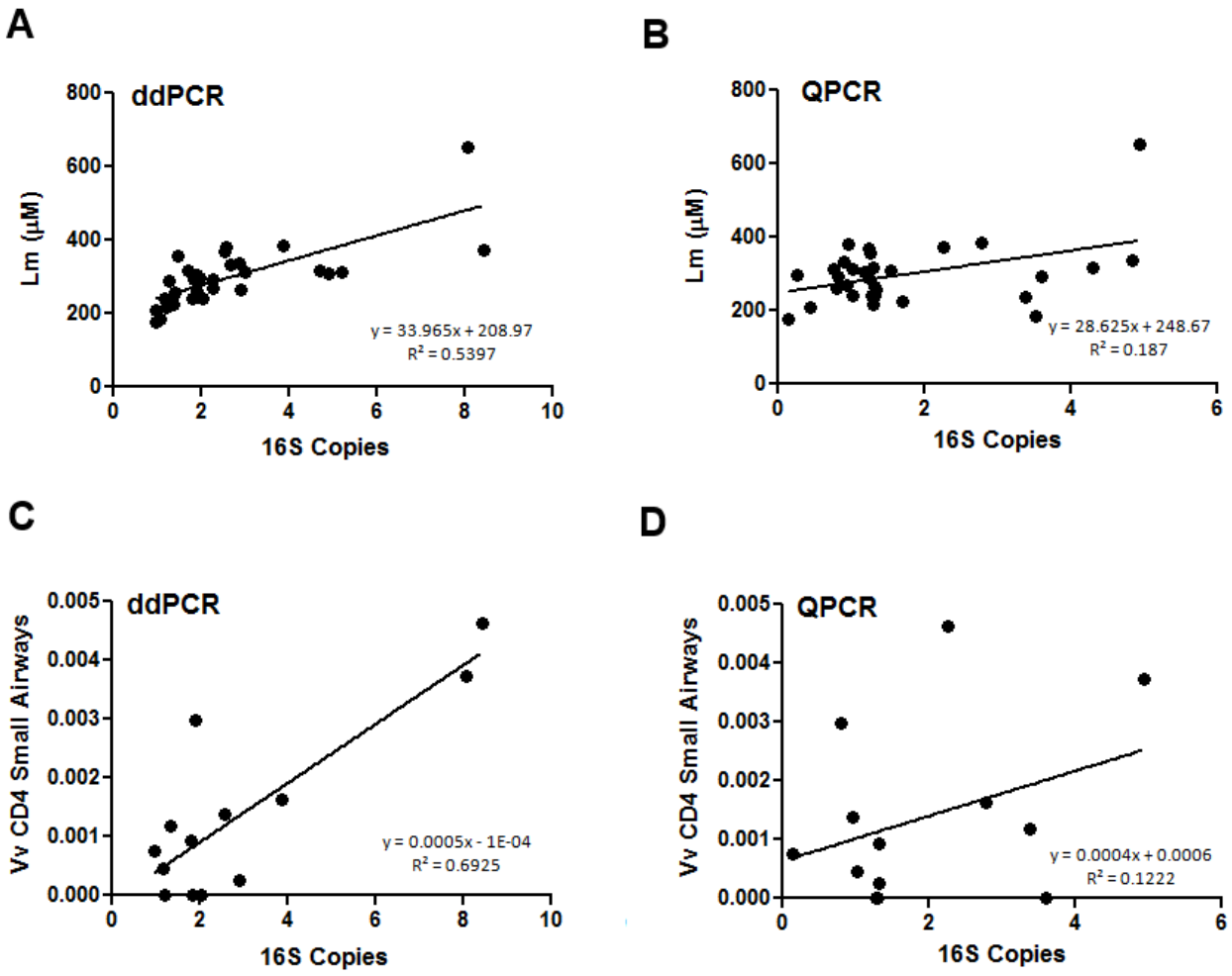


Figure 27: The Relationship between Quantitative Histology Parameters and ddPCR or qPCR Measurements. For all four panels linear regression analysis was used to test the correlation between the two variables.

5.4 Discussion

The first bacterial microbiome papers of the lungs were generated from materials obtained in bronchoalveolar lavage (BAL) and bronchial brushings (101,107). The total bacterial counts ranged from 10^3 to 10^5 total 16S within the lung (101,107,115,124). However, when similar assays were performed in resected lung tissue, these counts dropped to ranges between 1 and 10^2 total 16S per lung (109). The lower range of bacterial 16S impinges on the lower limit of detection for traditional qPCR assays and as such cannot be accurately quantified using this technique. In this study, we determined whether ddPCR significantly improves detection of bacterial load compared with traditional techniques of quantification. Compared with traditional qPCR, ddPCR has lower detection limits and a larger dynamic range of detection. Consistent with these properties, we found that the ddPCR assay reduced CV and thus the noise to signal ratio of bacterial detection, enabling robust quantification [188]. This is important because although there were no significant difference in total bacterial count in control and moderate COPD tissue samples [Figure 24], the ddPCR technique improved the tightness and dynamic range of the relationship between total bacterial count and important parameters of COPD such as Lm and CD4 counts in the small airways [Figure 27].

To date, most papers have not found a significant difference between the total bacterial load and COPD (101,108,109). However, there may be subtle but important differences in diversity of the bacterial microbiome between normal lungs and COPD lungs that might affect disease pathogenesis and progression (102,115). Our data suggest that using more sensitive PCR technology (ddPCR), we may gain important insights into potential disease mechanisms that may

have been elusive using the traditional qPCR approach. This approach would also be a way of validating or investigating specific bacterial species identified from unbiased sequencing and their potential role in COPD disease pathogenesis. ddPCR may be the preferred method given that many important OTUs (Operational Taxonomic Units) identified in these previous studies were found in very low relative abundance (102,107,109). The traditional qPCR technique may not have the ability to differentiate the samples owing to its relatively high detection limits. Another potential application of this technology may be in evaluating a select number of bacterial species in longitudinal studies in lieu of full Roche 454™ pyrotag or MiSeq™ Illumina sequencing, which are expensive. ddPCR may provide the needed sensitivity to follow specific low abundance bacterial species over time. The major advantage of ddPCR is in samples that contain relatively low abundance of bacterial load [Figure 26]. Moving the lung microbiome field beyond the cross-sectional experimental design (to longitudinal studies) has been one of the major limitations in discovering and confirming the important bacterial genera and species involved in the pathogenesis of the disease (207–209) and may provide the crucial technology needed to assess specific bacteria within the tertiary lymphoid follicles seen in very severe COPD.

This improvement may not be limited to the bacterial microbiome, ddPCR may be useful in detecting low copies of specific virus. ddPCR may also be able to analyze differences in bacterial strains and help to investigate the emergence of new strains (142,210) or and how they interact with the microbiome to help drive COPD progression. Overall this promising technology provides a measurable improvement over the traditional qPCR bacterial 16S assays used in assessing the bacterial load.

There were some limitations to the present study and ddPCR. We used a SYBR-green based assay rather than TaqMan-probe based methods for bacterial load quantification, which is thought to be more sensitive and less variable than SYBR-green based techniques (211). However, with strict standardization and optimization of procedures (as we did for the present study), these advantages of TaqMan-probe based assays over SYBR-green based methods largely disappear (212). Therefore, it is likely that ddPCR would be superior to TaqMan-based PCR with higher precision and faster throughput. A head-to-head comparison between these two methods would be needed to validate this hypothesis. A limitation associated with ddPCR is that it does not work well with high abundance samples and specifically for concentrations higher than ~ 10^5 target copies (202). This is due to the partitioning aspect of the technology, number of droplets generated, and the Poisson equation used to accurately measure the number of DNA copies in the samples (202). Along these lines the samples need to be diluted below 10^5 copies of the gene if ddPCR is to be used for target quantitation. Otherwise qPCR can be applied for measuring samples with high abundance. Additionally, using ddPCR, the sample processing time increases by approximately 45 minutes and depending on the number of samples the droplet read time adds 1-2 hours to the overall process. However, the increased time required to complete the assay may be an appropriate trade-off in low abundance samples because ddPCR is superior to the qPCR assay by allowing for extremely accurate quantification while reducing the overall 16S bacterial reads detected in the negative control background samples. ddPCR is a promising new technology that can potentially greatly advance the lung microbiome field by helping to move the field from hypothesis generating to hypothesis testing.

Chapter 6: Using Droplet Digital PCR to Elucidate the Role of *Haemophilus influenzae* in COPD

6.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is fast becoming a global health epidemic (2) and will be the 4th leading cause of death in the world (2). Previous studies have shown that active inflammation and remodeling of the airways occurs throughout the disease process (58,61,120,137,213). In a recent study by McDonough, et al. it was shown that the terminal bronchioles in COPD are significantly reduced before emphysematous destruction could be detected by MDCT (57).

An active adaptive immune response has been documented in COPD tissue (53). This adaptive response involves tertiary lymphoid follicles associated with the small airways and can be increasingly found as disease severity worsens (53). What drives this adaptive immune response is not well characterized. Some previous studies have shown that an autoimmune response to elastin may be present (141). Alternatively, bacteria, viruses, and other environmental factors could also be important for driving this response (71,78,142,144,210).

Recent studies on the bacterial microbiome in COPD have shown that there is a shift in bacterial community composition compared to controls (101,107,109). Additionally, a recent study has shown that there is a possible host response to bacteria within this shifted bacterial microbiome (214). *Haemophilus influenzae* was a bacterium predicted to be important between control and COPD GOLD 4 lung tissue samples (214). However, the bacterium was found to be associated

with control rather than GOLD 4 lung tissue and would seem to be at odds with the existing literature about the observation of *H.influenzae* and its role in exacerbations and generalized inflammation in COPD (142)(215)(216).

Using *H.influenzae* as a model bacterium, the data from previous studies suggests that it is possible specific bacteria identified in a shifted bacterial microbiome could provide targets for the adaptive immune response. This adaptive immune response could then in turn provide a continuous activation of the inflammation observed in COPD and provide a mechanism by which the terminal bronchioles may be remodeled and/or destroyed. Ultimately, this process could also eventually lead to a dramatic increase in tertiary lymphoid follicles over time.

The null hypothesis that was tested was that *H.influenzae* is not correlated with the adaptive immune response or its activation. There were two specific aims for this study. First, it investigated whether *H.influenzae* was associated with the adaptive immune system in mild and moderate COPD. Second, it investigated if there was a difference in the adaptive immune system activation between *H.influenzae* positive and negative tissue samples. The study involved two parts. The first part utilized "at risk" or control to GOLD 2 grade tissue samples to investigate whether *H.influenzae* could potentially illicit an immune response in mild and moderate COPD. The second part utilized multiple tissue samples from both control and GOLD 4 individuals. The multi-sampling was used to investigate disease within individuals and to test whether samples within the same individual showed different activation patterns when *H.influenzae* was present or not.

6.2 Methods

6.2.1 Sample Population and Tissue Procurement

Patient demographics from the mild and moderate study [Table 18] show a well matched population for sex, age, and smoking history. In contrast, demographic data for the adaptive immune activation arm show a higher number of males in controls versus the COPD GOLD 4 group [Table 19]. Tissue from the mild and moderate COPD group was obtained from patients undergoing lung resection therapy at St. Paul's hospital. Tissue resections were inflated with OCT, frozen in liquid nitrogen vapors, cored, and then stored in a -80ºC freezer for future use. Tissues from the adaptive immune activation group were from patients undergoing lung transplant or from donors for which no suitable recipient could be found. The lungs were inflated with air, frozen in liquid nitrogen vapors, sliced, cored, and then stored at -80ºC for future work. More information on either of these protocols can be found in previously published studies (53,57,173).

Table 18: Demographic Data of the Mild and Moderate COPD group

	Controls (n=28)	**GOLD 1** (n=21)	**GOLD 2** (n=25)
Age	65.7 ± 9.6	66.0 ± 8.9	63 ± 9.2
Sex (% Male)	57	58	68
Smoking History (cigarette-years)	895.4 ± 622.8	1061.8 ± 410.5	945.0 ± 555.7
FEV_1/FVC	77.4 ± 4.9	64.3 ± 4.3*	62.0 ± 7.0**
FEV_1 (percent predicted)	100.0 ± 12.5	89.9 ± 9.0†	69.0 ± 6.6**

* P<0.0001 between controls versus GOLD 1
**P<0.0001 between controls versus GOLD 2
†P<0.0001 between GOLD 1 versus GOLD 2

Table 19: Demographic Data of the Adaptive Immune Activation Group

	Controls (n=3)	**GOLD 4** (n=3)
Age	57.3 ± 5.7	59.3 ± 3.5
Sex (% Male)	100	33
FEV1/FVC	N/A	0.28 ± 0.06
FEV1 % Predicted	N/A	17.0 ± 6.2
Samples / Individual (n)	8 (3)	8(3)

6.2.2 Mild and Moderate COPD Group

DNA from two tissue samples per patient were extracted using the company defined protocol for the Qiagen DNeasy extraction kit (Qiagen Inc, Toronto, Canada). DNA concentration was quantified using a Nanodrop machine (Nanodrop products, Wilmington, USA). Each sample was quantified for *H.influenzae* using a modified assay for the QX200 ddPCR system (Bio-Rad Laboratories, Mississauga, Canada) (217,218). The samples were run in duplicate and if at least one of the duplicates were positive the sample was classified as positive for *H.influenzae* [Figure 28]. Sections from each tissue sample were mounted onto glass slides and stained for the following inflammatory cells: Macrophages, CD4+ T-cells, CD8+ T-cells, B-cells, and neutrophils. Point counting with a grid based system was used to obtain the volume fraction (Vv) of each inflammatory cell using Aperio (Leica Microsystems Inc., Concord, Canada) and

ImagePRO Plus software version 4.0 (MediaCybernetics Inc., Bethesda, MD). An H&E and movat pentachrome stain were completed in order to quantify total tissue percent, Vv of elastin, and Vv total collagen. Mean linear intercept (Lm) was measured on the H&E and movat sections with the average of the two being used for the overall Lm value.

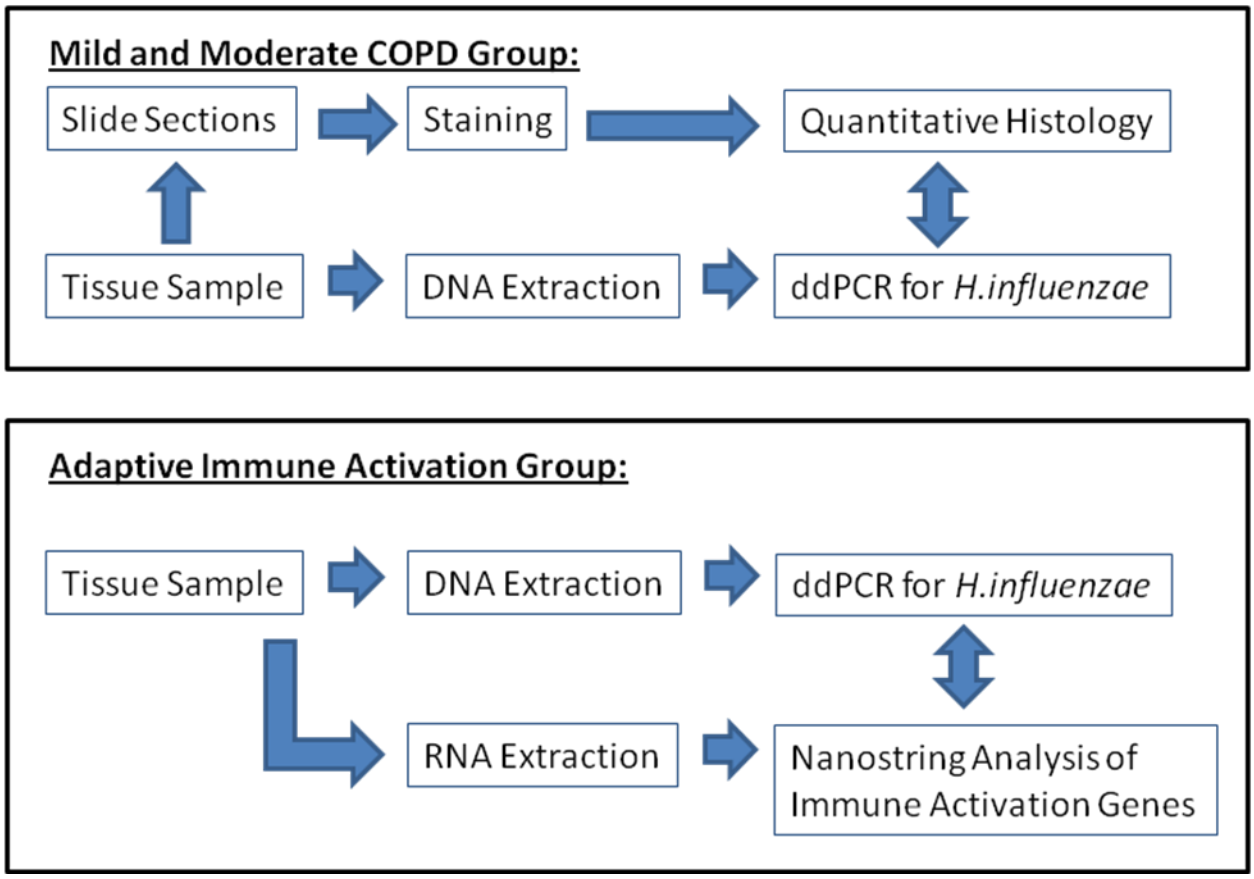


Figure 28: Overall Workflow of the Two Sample Groups used in this Study

6.2.3 Adaptive Immune Response Activation Group

The protocol was similar to the mild and moderate COPD cohort except that eight samples per patient were used instead of two [Figure 28]. The inflammatory cell quantitative histology was

expanded to include natural killer cells and eosinophils. The ddPCR protocol was used as previously specified. Elastin was quantified using an immunohistochemical stain instead of a Movat pentachrome stain. Collagen I and III were also measured in addition to total collagen. The Nanostring nCounter analysis system with a custom codeset [Table 20] (Nanostring Technologies, Seattle USA) was used to measure the adaptive immune response activation.

Table 20: List of Adaptive Immune Activation Genes used on Nanostring Platform.

CXCL13	BCAP	BTK	NFATC2
CD79A	NFATC3	CD19	TCRA
CD79B	ILF2	LYN	TCRB
CCR7	LAT2	BLNK	CD28
CR2	GRB2	CLNK	ZAP70
CD22	PTPN12	FCGR1A	LCK
FCRLA	TAL1	FCGR2A	FYN
BCL11A	PAG1	FCGR3A	FYB
NOPE	CD4	BCL6	LCP2
IGH@	CD8A	BCL10	LAT
IGLL1	IL17A	NFKB1	TRAF6
IGLL5	IL17F	CDC42	
CD45	FOXP3	SYK	

6.2.4 Statistical Analysis

In the mild and moderate COPD data set T-tests were used to determine whether *H.influenzae* positive tissue samples were significantly different versus *H.influenzae* negative tissue samples. ANOVA with Bonferroni correction was used to test for increases in Vv of inflammatory cells according to GOLD grade. Regression random forest analysis with Boruta feature selection (152) was used to analyze which quantitative histological measures were predictive of *H.influenzae* concentration. A multiple linear mixed-effect model was used to confirm whether these measurements were significant after correction for samples coming from the same patient. For the adaptive immune response activation analysis a regularized canonical correlation analysis (RCCA) was used to test if more genes were correlated with quantitative histological measures in the *H.influenzae* positive group versus negative group. Linear mixed-effect models with FDR correction were used as previously described (19) to test whether CD79a and TCRA gene expression correlated differently with the other adaptive immune activation genes in *H.infleunzae* positive and negative samples. A p-value of 0.05 and an FDR of less than 0.1 were considered significant. Statistical analysis was performed using R (Version 3.1.2), R Studio (Version 0.98.1091), and Prism v. 5 (GraphPad Software Inc. La Jolla California).

6.3 Results

6.3.1 *H.influenzae* Measurement with ddPCR

A comparison between qPCR and ddPCR was done for *H.infleunzae* measurement in tissue samples. Both assays targeted the protein D (hpd) gene of *H.influenzae*. For the qPCR a total of

2 patient's tissue were positive for *H.influenzae* (2.7%). For the ddPCR a total of 37 individuals were positive for *H.influenzae* (50%) [Table 21 & 22]. Further, the same 2 samples that were positive in the qPCR were also positive in the ddPCR. Additionally, every extraction negative control was negative for *H.influezae* as measured by ddPCR [Table 23]. From this data it was concluded that ddPCR was a more sensitive measure and for this study it was utilized for *H.influenzae* detection. The specificity of this assay in a qPCR system has previously been demonstrated (217).

Table 21: ddPCR on Each Tissue Sample for *H.influenzae*

	Positive Samples	Negative Samples	% Positive
At Risk	18	36	33.33
GOLD 1	11	29	27.50
GOLD 2	16	26	38.10

Table 22: ddPCR for Each Individual for *H.influenzae* Positivity

	Positive Individuals	Negative Individuals	% Positive
At Risk	15	13	53.57
GOLD 1	9	12	42.86
GOLD 2	13	12	52.00

Table 23: ddPCR for *H.influenzae* of Extraction Negative Controls run with the Tissue Samples

	Positive	Negative	% Positive
Run 1	0	14	0
Run 2	0	14	0
Run 3	0	14	0
Run 4	0	6	0

6.3.2 Mild and Moderate COPD Cohort

There was no difference between the Lm of controls (292.98 ± 77.15 µm), GOLD 1 (318.84 ± 69.35 µm), and GOLD 2 (304.31 ± 65.27 µm) groups ($P > 0.05$). When all tissue samples were examined, regardless of GOLD grade, the *H.infleunzae* positive were different from their *H.influenzae* negative counterparts by having lower Vv of elastin and higher Vv of macrophages (P-value < 0.02). However, with Bonferroni correction for multiple comparisons these were no longer significant [Figure 29]. When the data was stratified by GOLD grade, CD68+ macrophages were significantly increased in GOLD 1 and GOLD 2 *H.influenzae* positive samples compared to negative samples ($P < 0.05$) after Bonferroni correction [Figure 30A]. Interestingly, CD79a+ B-cells were significantly increased in the GOLD 2 *H.influenzae* positive group compared to the *H.influenzae* negative group (P-value < 0.05) after Bonferroni correction [Figure 30B]. Boruta feature selection with a regression random forest analysis identified the Vv of CD79a+ B-cells and CD68+ macrophages as the most important predictors of determining the *H.influenzae* concentration within lung tissue samples. The Vv of elastin was a tentative marker

of *H.influenzae* concentration. In other words, an equal number of votes for importance and unimportance of the Vv of elastin to predict *H.influenzae* was made by the random forest classifier. All other measurements were determined to be unimportant for prediction. Together the measurements predicted 14.2% of the observed variation of *H.influenzae* according to Random Forest. Using a multiple linear mixed-effect model analysis to correct for tissue samples being obtained from the same patient both the Vv CD68+ macrophages and CD79a+ B-cells remained statistically significant (P-value < 0.002) [Table 24].

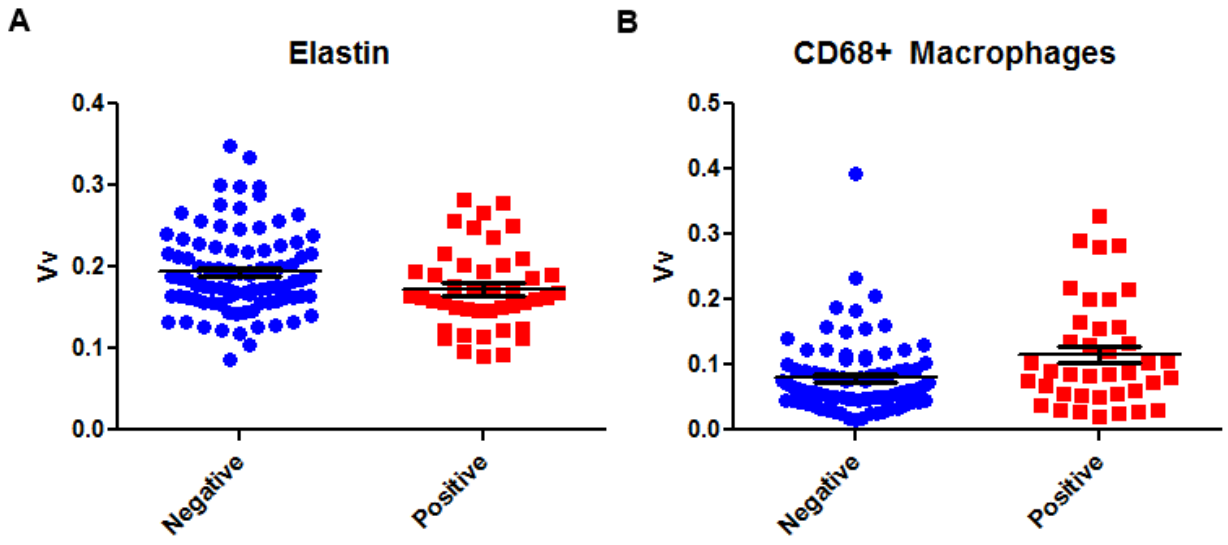


Figure 29: Significant Volume Fraction Differences between *Haempohilus influenzae* Positive and Negative Tissue Samples A) Vv of Elastin based on H.influenzae positivity. The P-value was less than 0.02 for student T-test but with Bonferroni correction P > 0.05. B) Vv of CD68+ macrophages based on H.influenzae positivity. The P-value was less than 0.02 for student T-test but with Bonferroni correction P-value > 0.05.

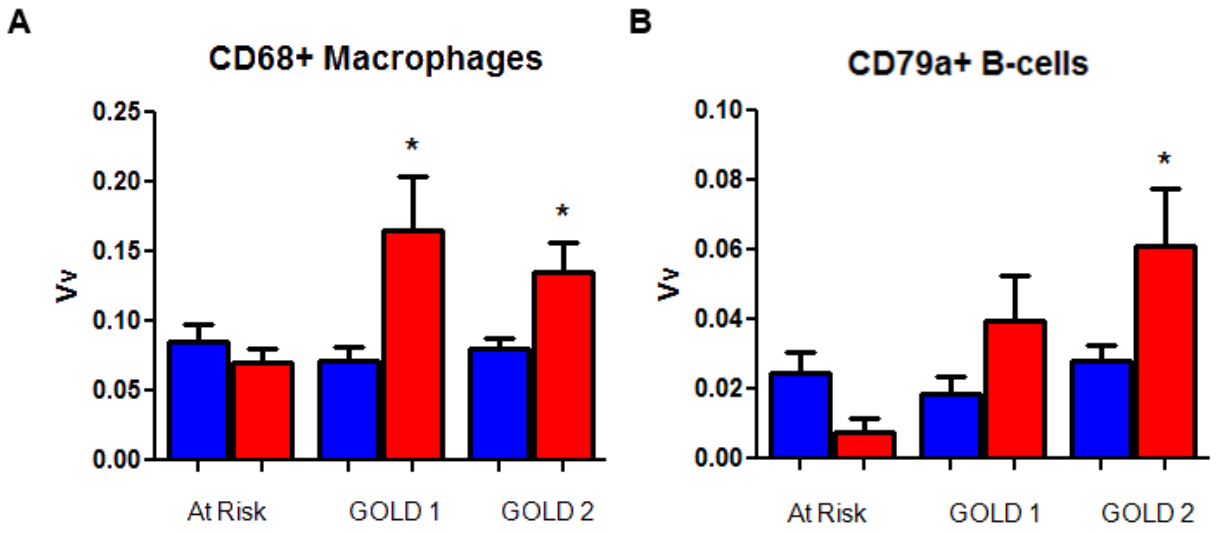


Figure 30: Significant Differences between *Haemophilus influenzae* Positive and Negative Tissue Samples Stratified by GOLD Grade. A) Vv of CD68+ macrophages by GOLD grade and H.influenzae positivity. A significant difference between GOLD 1 and GOLD 2 grade H.influenzae positive and negative samples was found (*P-value < 0.05 with Bonferroni correction). B) Vv of CD79a+ B-cell by GOLD grade and H.influenzae positivity. A significant difference between GOLD 2 grade H.influezae positive and negative samples was observed (*P-value < 0.05 with Bonferroni correction). Blue bars represent *H.influenzae* negative samples while the red bars represent *H.influenzae* positive samples.

Table 24: Multivariate Linear Mixed Effect Analysis of *H.influenzae* Concentration and Quantitative Histology Meaurements

Measurement	Value	Std. Error	DF	T-value	P-value
Tissue Percent	0.0003	0.0001	15	2.10	0.05
PMN (Alv)	-0.0236	0.0383	15	-0.61	0.55
Lm	0.00002	0.00001	15	1.29	0.22
Macrophages (Alv)	**0.0626**	**0.0159**	**15**	**3.94**	**0.001**
Elastin (Alv)	-0.02	0.0144	15	-1.38	0.19
CD8 (Alv)	-0.0152	0.0207	15	-0.73	0.48
CD4 (Alv)	-0.0275	0.0209	15	-1.31	0.21
B cell (Alv)	**0.0786**	**0.0194**	**15**	**4.05**	**0.001**

6.3.3 Adaptive Immune Response Activation Cohort

Regularized Canonical Correlation analysis (RCCA) was used to explore the relationships between Vv of inflammatory immune cells and adaptive immune activation genes in *H.influenzae* negative and positive samples. Using RCCA *H.influenzae* negative samples had less positive correlations with adaptive immune response genes than *H.influenzae* positive samples with respect to Vv of CD68+ macrophages, Vv CD4+ T-cells, Vv of eosinophils, Vv of CD79a+ B-cells, and CD8+ T-cells [Figure 31A & B]. Using network analysis and an R-value cutoff of 0.7 a similar pattern can be observed between *H.influenzae* negative samples and *H.influenzae* positive samples [Figure 31C & D]. For the *H.influenzae* negative samples most

correlations were related to a negative correlation between Vv of neutrophils and adaptive immune response genes [Figure 31C]. In contrast, *H.influenzae* positive samples had strong positive correlations between Vv of CD68+ Macrophages, Vv of CD79a+ B-cells, and Vv of CD4+ T-cells and adaptive immune response genes [Figure 31D]. However, even in the *H.influenzae* positive samples the negative correlation between Vv of neutrophils and adaptive immune response genes was still present [Figure 31D].

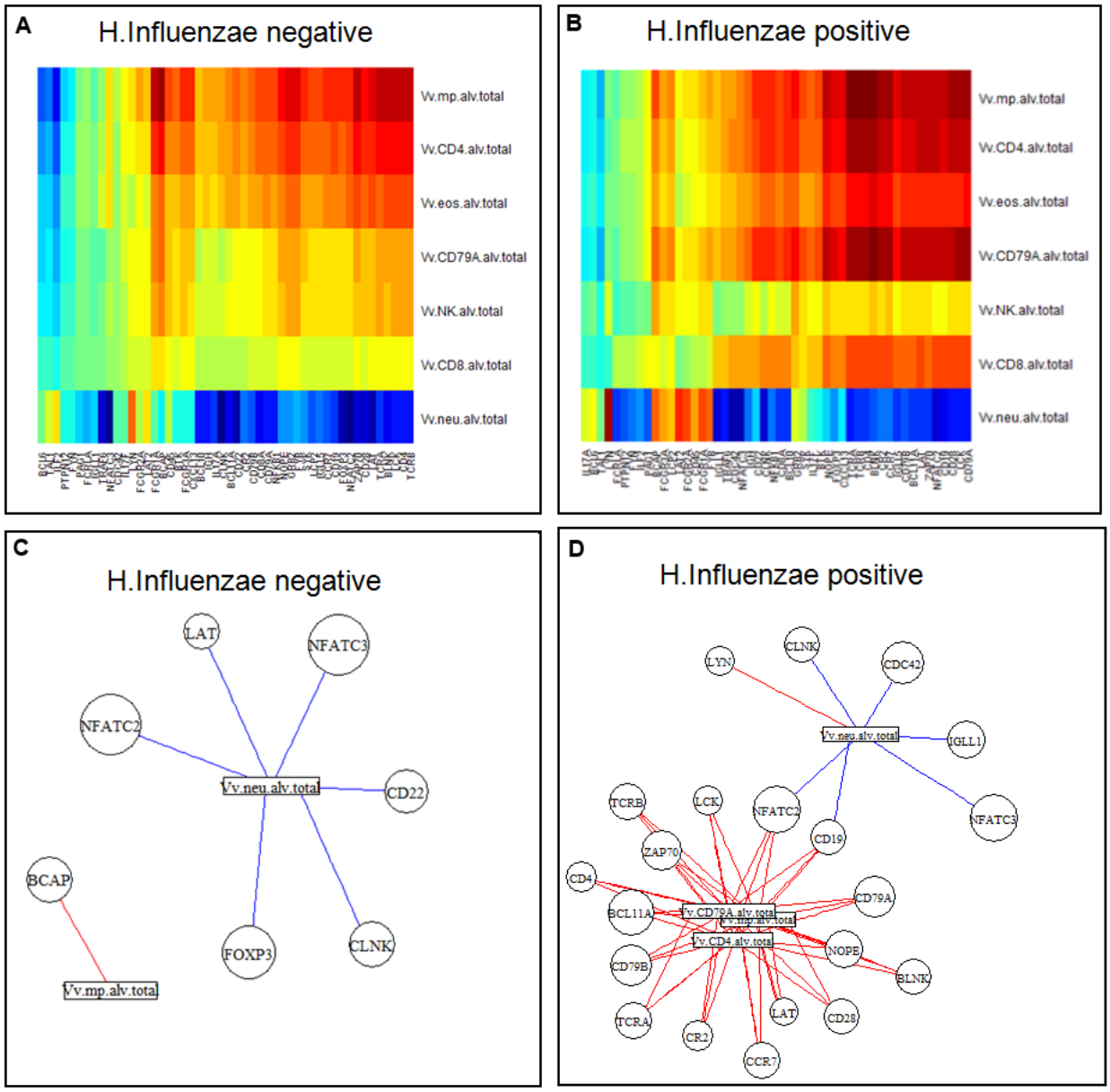


Figure 31: Regularized Canonical Correlation Analysis of Adaptive Immune System Activation Genes and Volume Fraction of Inflammatory Cells for *Haemophilus influenzae* Positive or Negative Samples A) Heatmap of Regularized Canonical Correlation (RCC) analysis of H.influenzae negative tissue samples (n=28) of quantitative histology analysis of Vv of inflammatory cells versus adaptive immune system activation genes. B) Heatmap of RCC analysis of H.influenzae positive tissue samples (n=12) of quantitative histology analysis of Vv of inflammatory cells versus adaptive immune system activation genes. For the heatmap stronger positive correlations are represented by darker red, stronger negative correlations are represented by darker blue, and no correlation is represented by a green-yellow. C) Network analysis of H.influenzae negative tissue samples (n=28) between Vv of inflammatory cells and adaptive immune system activation genes. D) Network analysis of H.influenzae positive tissue samples

(n=12) between Vv of inflammatory cells and adaptive immune system activation genes. For the network analysis an R-value cutoff of 0.7 was used to create the network.

When analyzing the CD79a gene expression correlation to other adaptive immune response gene expression, *H.influenzae* negative samples had more correlations survive FDR correction than *H.influenzae* positive samples [Table 25 & 26]. However, some notable differences between the two groups exist. The *H.influenzae* positive group had a significant positive correlation between CD79a and IGLL5, BCL10, and MEF2C that was not seen in the *H.influenzae* negative group. There were also significant negative correlations between CD79a and FYN and CD22 that was not observed in the *H.influenze* negative group. This same pattern held for the TCRA gene expression comparison to adaptive immune response genes in *H.influenzae* negative and positive samples [Table 27 & 28]. Similar to the CD79a comparison the *H.influenzae* group had significant correlations to TCRA to BCL10, IGLL5, and NFATC3 not seen in the *H.influenzae* group. No significant negative correlation difference was seen between the two groups with respect to TCRA comparisons.

Table 25: Significant CD79a Gene Comparison to Adaptive Immune Response Genes in *H.influenzae* Negative Samples

Comparison	Coefficient	T-stat	P-value	FDR
CD79A_CD19	0.780799	10.52682	7.20E-12	8.63E-10
CD79A_BLNK	1.900204	10.14001	7.77E-09	3.11E-07
CD79A_CCR7	1.014037	6.93754	1.58E-07	3.16E-06
CD79A_CR2	0.551909	6.141866	3.89E-07	4.67E-06
CD79A_CD22	1.169934	6.280659	3.76E-07	4.67E-06
CD79A_FOXP3	1.422133	6.223797	3.65E-07	4.67E-06
CD79A_BCL11A	1.261514	6.277124	7.04E-07	7.56E-06
CD79A_CD28	1.2482	5.847284	9.01E-07	7.88E-06
CD79A_IgD	0.783689	5.896684	9.19E-07	7.88E-06
CD79A_TCRB	1.298712	7.335565	1.62E-06	1.23E-05
CD79A_NFATC2	1.370311	6.2637	3.85E-06	2.71E-05
CD79A_LCK	1.549997	6.647797	4.50E-06	3.00E-05
CD79A_ZAP70	1.139924	5.942258	8.61E-06	4.92E-05
CD79A_LAT	1.173926	5.233922	1.09E-05	5.96E-05
CD79A_CXCL13	0.364781	4.918662	1.41E-05	7.06E-05
CD79A_TCRA	1.059553	5.715847	4.14E-05	0.000191
CD79A_IGH	0.6416	3.733348	4.35E-04	0.00158
CD79A_CLNK	0.866986	3.777744	5.27E-04	0.001861
CD79A_CD8A	1.231034	3.249233	1.69E-03	0.005192
CD79A_FYB	0.843453	2.795239	6.00E-03	0.017549
CD79A_MEF2C	1.966291	2.794445	8.67E-03	0.024781
CD79A_BCAP	1.044198	2.71695	9.18E-03	0.025027
CD79A_CD4	1.395452	3.009078	1.02E-02	0.027104
CD79A_KLHL6	0.891159	2.53842	1.17E-02	0.030508
CD79A_NFATC3	1.755937	2.341437	1.88E-02	0.045471
CD79A_LCP2	1.113228	2.338178	1.89E-02	0.045471
CD79A_SYK	1.201478	2.186159	2.69E-02	0.058754
CD79A_FCGR1A	0.732889	2.113526	3.21E-02	0.068743
CD79A_FCRLA	0.657863	2.019196	4.08E-02	0.080975
CD79A_BCL10	1.79148	2.019797	4.14E-02	0.080975
CD79A_IL17A	0.537263	2.005341	4.18E-02	0.080975
CD79A_TRAF6	1.297757	1.997486	4.31E-02	0.082138

Table 26: Significant CD79a Gene Comparison to Adaptive Immune Response Genes in *H.influenzae* Positive Samples

Comparison	Coefficient	T-stat	P-value	FDR
CD79A_CCR7	1.506857	9.01034	1.63E-05	0.000392
CD79A_NFATC2	1.323612	9.812951	6.17E-06	0.000392
CD79A_ZAP70	1.177633	9.053367	1.55E-05	0.000392
CD79A_LAT	1.229212	9.262671	1.19E-05	0.000392
CD79A_TCRA	1.094349	6.566869	2.38E-05	0.000476
CD79A_CD19	0.818628	6.225925	2.96E-05	0.000508
CD79A_BLNK	1.612251	5.345454	1.56E-04	0.001336
CD79A_CD38	0.851346	4.502223	3.78E-04	0.002688
CD79A_BCL11A	1.316035	6.414505	6.84E-04	0.003729
CD79A_TCRB	1.307519	6.438638	6.57E-04	0.003729
CD79A_CD4	2.163954	5.424618	9.88E-04	0.005115
CD79A_LCK	1.548863	6.17395	1.02E-03	0.005115
CD79A_BCL10	3.640939	3.319375	1.95E-03	0.009143
CD79A_MEF2C	2.838889	3.281242	2.13E-03	0.00921
CD79A_KLHL6	1.328388	4.272185	6.10E-03	0.023624
CD79A_CD8A	1.305069	2.715739	7.96E-03	0.02985
CD79A_CLNK	1.087249	4.93844	9.90E-03	0.035646
CD79A_FOXP3	1.278399	4.833721	1.22E-02	0.038589
CD79A_CR2	0.848399	3.037497	1.32E-02	0.039522
CD79A_IGH	0.75281	2.481355	1.35E-02	0.039522
CD79A_IGLL5	2.811344	4.784527	1.35E-02	0.039522
CD79A_CXCL13	0.394369	2.532313	1.55E-02	0.044251
CD79A_CD28	1.092285	3.034235	1.79E-02	0.049848
CD79A_CD83	1.016485	3.150857	2.09E-02	0.053388
CD79A_NFKB2	1.684044	3.258202	2.76E-02	0.067488

Table 27: Significant TCRA Gene Comparison to Adaptive Immune Response Genes in *H.influenzae* Negative Samples

Comparison	Coefficient	T-stat	P-value	FDR
TCRA_LCK	1.19819	11.98047	3.90E-13	2.34E-11
TCRA_CD4	1.667912	8.145546	2.00E-08	2.67E-07
TCRA_CD28	0.931163	6.832731	6.17E-08	6.73E-07
TCRA_BLNK	1.311301	8.4106	1.60E-07	1.60E-06
TCRA_FYB	0.857293	5.753182	1.43E-06	1.07E-05
TCRA_LAT	0.894218	6.470268	1.69E-06	1.19E-05
TCRA_CCR7	0.639872	5.726452	2.26E-06	1.29E-05
TCRA_ZAP70	0.805259	6.132637	2.11E-06	1.29E-05
TCRA_CXCL13	0.253755	5.384682	4.44E-06	2.05E-05
TCRA_CD45	0.879795	5.237601	1.12E-05	4.48E-05
TCRA_SYK	1.416838	5.071943	1.31E-05	5.03E-05
TCRA_CD8A	1.100547	4.858444	1.65E-05	6.02E-05
TCRA_BCAP	0.99976	4.819555	1.87E-05	6.59E-05
TCRA_LCP2	1.194699	4.553476	4.27E-05	0.000138
TCRA_CD86	0.978912	4.406488	6.31E-05	0.000199
TCRA_CD79A	0.527148	5.545141	6.89E-05	0.000212
TCRA_FCGR1A	0.829934	4.284347	1.08E-04	0.000316
TCRA_GRB2	1.837488	3.962258	2.21E-04	0.00059
TCRA_CLNK	0.64784	4.221627	2.32E-04	0.000605
TCRA_FOXP3	0.928439	4.781123	2.46E-04	0.000628
TCRA_IGLL5	0.97839	5.740493	2.68E-04	0.000671
TCRA_NFATC2	1.035511	7.2031	3.15E-04	0.000771
TCRA_CD79B	0.58734	3.902144	3.38E-04	0.00081
TCRA_CD22	0.637453	3.748444	4.10E-04	0.000965
TCRA_IGH	0.447259	3.728709	4.34E-04	0.001001
TCRA_CD19	0.361368	3.734579	5.75E-04	0.001302
TCRA_NOPE	0.929116	3.69054	1.61E-03	0.003444
TCRA_CR2	0.26617	3.205332	1.92E-03	0.004037
TCRA_BTK	0.833317	3.158054	2.54E-03	0.005164
TCRA_KLHL6	0.713602	3.057987	3.01E-03	0.00602
TCRA_MEF2C	1.718349	3.786635	6.40E-03	0.012004
TCRA_NFKB1	1.653795	2.719885	8.54E-03	0.015716
TCRA_PAG1	0.833182	2.676108	9.58E-03	0.016912
TCRA_FCGR3A	0.543658	2.415489	1.56E-02	0.025973
TCRA_CD83	0.44248	2.457512	1.81E-02	0.02941
TCRA_LAT2	0.601486	2.145731	3.33E-02	0.051246
TCRA_LYN	1.173777	2.146468	3.73E-02	0.056583

Comparison	Coefficient	T-stat	P-value	FDR
TCRA_IgD	0.271374	1.977595	4.58E-02	0.067859
TCRA_NFKB2	0.695319	2.072164	5.29E-02	0.077422
TCRA_ILF2	-2.35853	-1.86018	5.90E-02	0.085371

Table 28: Significant TCRA Gene Comparison to Adaptive Immune Response Genes in *H.influenzae* Positive Samples

Comparison	Coefficient	T-stat	P-value	FDR
TCRA_LCK	0.841584	10.09623	2.70E-07	2.94E-05
TCRA_LAT	1.138591	8.094045	8.72E-07	3.07E-05
TCRA_ZAP70	0.937201	9.534567	4.84E-06	8.30E-05
TCRA_CD4	1.881754	9.192716	7.33E-06	0.00011
TCRA_CD79A	0.746692	8.488665	1.80E-05	0.00024
TCRA_BLNK	1.403876	6.596051	2.10E-05	0.000252
TCRA_CD79B	1.170236	7.260623	1.01E-04	0.000932
TCRA_CCR7	1.16318	7.292309	9.63E-05	0.000932
TCRA_CD28	1.205559	7.310173	9.38E-05	0.000932
TCRA_NFATC2	1.010024	7.115317	1.26E-04	0.001077
TCRA_CD8A	1.257037	4.388448	2.14E-04	0.001635
TCRA_CD19	0.692503	5.882776	9.36E-04	0.004883
TCRA_IGH	0.80751	3.734692	1.21E-03	0.006051
TCRA_FOXP3	1.037762	5.316675	2.61E-03	0.01158
TCRA_KLHL6	1.05293	5.137971	3.65E-03	0.015123
TCRA_IGLL5	2.217046	4.714069	8.42E-03	0.030628
TCRA_CLNK	0.843214	4.588025	1.09E-02	0.037361
TCRA_MEF2C	2.030073	2.527369	1.13E-02	0.037657
TCRA_BTK	1.099538	2.6329	1.18E-02	0.038235
TCRA_CR2	0.863781	3.841628	1.27E-02	0.039223
TCRA_CD83	0.805702	2.732233	1.51E-02	0.044096
TCRA_BCL10	2.439985	2.271143	2.05E-02	0.055814
TCRA_CD22	-0.72712	-3.779	2.93E-02	0.073224
TCRA_NFATC3	1.757153	2.206221	2.87E-02	0.073224
TCRA_TRAF6	1.308358	2.086464	3.68E-02	0.09003
TCRA_Fas	1.271987	1.940404	4.15E-02	0.099622

6.4 Discussion

This study builds upon previous literature on *H.influenzae* and COPD(215)(182,216). However, most of these studies were performed in sputum cultures and this is one of the first studies to investigate *H.influenzae* in lung tissue specifically. One recent study by King, et al. investigated

COPD lung tissue to specifically identify response differences in T-cells to *H.influenzae* between COPD, non-smoking, and smoking control tissue (219). Interestingly, they found an increased production in certain cytokines in COPD T-cells stimulated with *H.influenzae* (219). Their study focused on two specific cell types while our study tries to investigate the adaptive immune system response to *H.influenzae* as a whole.

As compared to previous studies the overall *H.influenzae* load in lung tissue was quite low and qPCR was not as sensitive as ddPCR in identifying tissue samples with *H.influenzae*. The results show that before any noticeable emphysematous destruction can be measured in lung tissue *H.influenzae* positive samples have increased Vv of both macrophages and B-cells dependent on GOLD grade [Figure 30]. Further, a larger number of correlations between quantitative histology and adaptive immune activation genes can be found in *H.influenzae* positive tissue samples [Figure 31]. Finally, specific differences in the types of gene to gene correlations with either CD79a or TCRA could be found between *H.influenzae* positive and negative samples. Most notably IGLL5 and BCL10 were positively correlated with both CD79a and TCRA in the *H.influenzae* positive samples but not in negative samples. Overall this data provides evidence that an adaptive immune response could be directed towards *H.influenzae* in lung tissue and that this may happen before emphysema is detectable.

H.influenzae has been known to be a common colonizer of the nasopharynx along with other species such as *Streptococcus pneumoniae* (180). The dynamics between bacteria in the nasopharynx is complex with some data showing that *H.influenzae* can increase clearance of *S.pneumoniae* by activating neutrophils and complement-dependent clearance pathways (180). It

is possible that in healthy individuals an active immune response to this bacterium is not normally made. However, in lung diseases, such as COPD, impaired mucociliary clearance (129), mucus plugging (220), and other immune impairments (221) could lead to the bacterium travelling into the lower airways, staying in these locations, and eliciting an immune response. If the immune response in COPD is directed against what are normally commensal organisms then it could explain why vaccination of *H.influenzae* has been unsuccessful in lowering overall bacterial loads (222). It could also support the notion that the emergence of new strains of the bacterium in COPD (142) could be due to it trying to survive in an environment that has become increasingly hostile. Some inflammatory data shows that dendritic cells can be activated by *H.influenzae* and the bacterium can cause significantly higher release of IL-23, IL-12p70, and IL-10 cytokines (132). An increase in the anti-inflammatory IL-10 could support the idea that under normal conditions this bacterium does not elicit the strong adaptive immune response seen in COPD.

Alternatively, *H.influenzae* could be needed to initially drive the inflammation in COPD. Although Teo, et al. showed no decrease in bacterial load there was a significant decrease in antibiotic usage and a potential increase in quality of life in the *H.influenzae* vaccinated groups in their meta analysis (222). Other literature shows that after viral infection there is outgrowth of *H.influenzae* in COPD patients (223) and worsening of daily symptoms in those with *H.influenzae* (224). A recent study has shown that *H.influenzae* along with cigarette smoke can better reproduce a human like COPD phenotype in mice (215). In fact it was found that low dose of *H.influenzae* created a macrophage dominated inflammatory profile (215) and this

observation matches well with previous literature on COPD in lung tissue (53). Further, without bacterial exposure goblet cell metaplasia and lymphoid aggregates were not observed (215).

Our data neither confirms nor disproves either of these two possible scenarios. Both of these hypotheses stipulate an adaptive immune response to *H.influenzae* which our data supports. However, further work needs to be done to show whether it is the bacterium that drives the inflammation or if the targeting of the bacterium is due to a dysfunctional, over active inflammatory response caused by other factors. Some data supports the idea that it could be the specific individual's response to the bacterium that can affect inflammatory cytokine production and not the bacterium itself (225). These variations from person to person could ultimately dictate why some individuals progress in GOLD grade while others do not. However, other data support the idea that *H.influenzae* has a direct role in causing the progression of COPD. For example, one such study found that *H.influenzae* lowered IgA specific antibodies targeted against it and increased MMP-9 in sputum samples (226). It is likely that both hypothesis are true and that a unique interplay between the deficiencies in some individuals host response and *H.influenzae* are at play. In other words, smoking and *H.influenzae* alone are not enough to drive disease since only a small percentage of smokers develop COPD. Specific differences in host immune response to the bacterium combined with the inflammatory environment of smoking are most likely the key driver of the observed correlations we report here on the adaptive immune response and *H.influenzae*.

Although this study builds upon the existing ground work of literature on *H.influenzae* and COPD there are a few caveats. First, secondary studies in lung tissue, especially with respect to

the adaptive immune response activation, need to be completed due to the small numbers of total patients analyzed. Second, the findings in this study are correlations and more in-depth analysis of specific cell types, most notably B-cells, CD4+ T cells, and macrophages, need to be investigated. By examining how the response to *H.influenzae* differs between COPD and smoking controls it may be possible to figure out what specific part of the immune response is dysfunctional in those that rapidly progress to very severe COPD. Third, although there are increased correlations in the *H.influenzae* group between gene expression and Vv of the specific cells they may not be necessarily activated. As an example, correlations to specific genes such as IL-17A, IL-17F, and CXCL13 that can be more indicative of immune cell activation (227,228) are absent in the *H.influenzae* positive group [Figure 31]. This could mean that *H.influenzae* supports a generalized increase in CD4+ T-cells as well as B-cells without the corresponding or adequate activation of these respective cells.

Although the ddPCR technology gets us tantalizingly close to identifying a potential mechanism by which *H.influenzae* could impact COPD severity a few other studies will need to be completed before this can be confirmed. First, it will need to be confirmed either in a repeat experiment or one with a slightly different protocol whether or not immune activation genes along with general immune genes are correlated with *H.influenzae.* If it is just a general increase then investigating why there is no subsequent activation of these specific cells would be a highly relevant area to pursue. In contrast, if activation can be shown by adding more specific genes that are markers for immune activation then investigating the tertiary lymphoid follicles for *H.infleunzae* may eventually be able open a window into a potential target of the adaptive immune system in COPD.

Chapter 7: Conclusion

This research successfully builds upon previous studies on the bacterial microbiome by showing that changes in the bacterial community can be observed as early as moderate (GOLD2) COPD. Secondly, they show that these changes in the bacterial community correlated with both the increase in the inflammatory response and with the structural changes associated with tissue repair. Third, the results obtained indentify a list of OTUs that are potentially important for driving some of these changes. Fourth, *Haemophilus influenzae,* a bacterium identified in this list of OTUs, correlated with the volume fraction of B-cell and macrophages infiltrating into the tissue before emphysematous tissue destruction appeared. Fifth, *Haemophilus influenzae* can induce small differences in adaptive immune cell activation. Taken together this data strongly suggests that the pathogenesis of COPD may be influenced by specific bacteria located within the microbiome of the lungs.

The data reviewed in Chapter 1and previously reported as part of my MSc thesis of which most is previously published (109) suggested that the Firmicute phyla contributed to the severity of COPD. In contrast the data presented in Chapter 3indicate Proteobacteria may be more important in mild and moderate COPD. Moreover this opinion was reinforced in Chapter 4 where the data showed that OTUs in the Proteobacteria phyla correlate with inflammatory markers and structural changes associated with the host response to injury and tissue repair. Although these findings differ from the previous literature on the bacterial microbiome in COPD (109), other data also suggests that Proteobacteria are important in the pathogenesis of COPD (101). One potential reason for this discrepancy is the fact that the new data presented in

Chapter 3 and Chapter 4 provide information about parameters of the host response that were not investigated in previous studies showing that Firmicutes were important. In fact the major difference between the two studies is that new work presented in Chapter 3 and 4 on the response to the bacterial microbiome was measured in terms of the specific features of the host response to injury rather than to estimates of the decline in lung function made from the GOLD category of COPD. Additionally differences between the findings presented in Chapter 4 and Chapter 3 are understandable since comparisons were made between tissue obtained from patients with end stage COPD treated by lung transplantation and from patients with very severe COPD to control lungs obtained from donors in Chapter 4 while in Chapter 3 tissue was obtained for the data from patients with mild to moderate COPD that required lung resection as treatment for lung cancer. A second possible reason for this difference is that the difference was based on the phenomenon originally described by Huang, et al. (108) where COPD GOLD 4 had two different groups with respect to the predominant phyla, one that was Firmicutes and the other that was Proteobacteria. In that the samples were obtained from a small group of patients that were biased towards e Firmicute in the data reported in Chapter 1 and Proteobacteria in the larger group of cases from different levels of COPD in Chapter 4.

Future studies based on larger groups of subjects with better sampling techniques and more exhaustive analytical procedures will be able to resolve these differences and improve our understanding of the nature of the bacterial microbiome as well as the factors that control the host response to it. Overall, the culmination of all the work presented in this thesis provides a good first step in providing support for the central hypothesis that the bacterial microbiome in lung tissue does have a role to play in the pathogenesis of COPD.

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec 15;380(9859):2095–128.

2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS medicine. 2006 Nov;3(11):e442.

3. Organization TWH. The global burden of disease: 2004 update. Geneva, Switzerland: WHO Press; 2008.

4. Gallus S, Lugo A, La Vecchia C, Boffetta P, Chaloupka FJ, Colombo P, et al. Pricing Policies And Control of Tobacco in Europe (PPACTE) project: cross-national comparison of smoking prevalence in 18 European countries. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation. 2014 May;23(3):177–85.

5. Agaku IT, King BA, Dube SR, Centers for Disease C, Prevention. Current cigarette smoking among adults - United States, 2005-2012. MMWR Morbidity and mortality weekly report. 2014 Jan 17;63(2):29–34.

6. Levy D, Rodriguez-Buno RL, Hu TW, Moran AE. The potential effects of tobacco control in China: projections from the China SimSmoke simulation model. Bmj. 2014;348:g1134.

7. Baty F, Putora PM, Isenring B, Blum T, Brutsche M. Comorbidities and burden of COPD: a population based case-control study. PloS one. 2013;8(5):e63285.

8. Zar HJ, Ferkol TW. The global burden of respiratory disease-Impact on child health. Pediatric pulmonology. 2014 May;49(5):430–4.

9. Kim J, Rhee CK, Yoo KH, Kim YS, Lee SW, Park YB, et al. The health care burden of high grade chronic obstructive pulmonary disease in Korea: analysis of the Korean Health Insurance Review and Assessment Service data. International journal of chronic obstructive pulmonary disease. 2013;8:561–8.

10. Punekar YS, Shukla A, Mullerova H. COPD management costs according to the frequency of COPD exacerbations in UK primary care. International journal of chronic obstructive pulmonary disease. 2014;9:65–73.

11. Pasquale MK, Sun SX, Song F, Hartnett HJ, Stemkowski SA. Impact of exacerbations on health care cost and resource utilization in chronic obstructive pulmonary disease patients with chronic bronchitis from a predominantly Medicare population. International journal of chronic obstructive pulmonary disease. 2012;7:757–64.

12. Lamprecht B, McBurnie MA, Vollmer WM, Gudmundsson G, Welte T, Nizankowska-Mogilnicka E, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. Chest. 2011 Apr;139(4):752–63.

13. Yoon HI, Sin DD. Confronting the colossal crisis of COPD in China. Chest. 2011 Apr;139(4):735–6.

14. Agustin C, Alison L, Agustina M, Demian G, Silvana C, Edgardo S. The Epidemiology and Burden of COPD in Latin America and the Caribbean: Systematic Review and Meta-Analysis. Copd [Internet]. 2013 Oct 10; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24111903

15. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clinics in chest medicine. 2014 Mar;35(1):7–16.

16. Guarnieri MJ, Diaz JV, Basu C, Diaz A, Pope D, Smith KR, et al. Effects of woodsmoke exposure on airway inflammation in rural guatemalan women. PloS one. 2014;9(3):e88455.

17. Liu Y, Lee K, Perez-Padilla R, Hudson NL, Mannino DM. Outdoor and indoor air pollution and COPD-related diseases in high- and low-income countries. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2008 Feb;12(2):115–27.

18. Smith M, Li L, Augustyn M, Kurmi O, Chen J, Collins R, et al. Prevalence and correlates of airflow obstruction in ~317 000 never-smokers in China. The European respiratory journal [Internet]. 2014 Mar 13; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24603814

19. Campbell JD, McDonough JE, Zeskind JE, Hackett TL, Pechkovsky DV, Brandsma CA, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. Genome medicine. 2012 Aug 31;4(8):67.

20. De Serres F, Blanco I. Role of alpha-1 antitrypsin in human health and disease. Journal of internal medicine [Internet]. 2014 Mar 24; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24661570

21. Alam S, Li Z, Atkinson C, Jonigk D, Janciauskiene S, Mahadeva R. Z alpha-1 antitrypsin confers a pro-inflammatory phenotype that contributes to COPD. American journal of respiratory and critical care medicine [Internet]. 2014 Mar 4; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24592811

22. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS, Committee GS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Llung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med. 2001;163(5):1256–76.

23. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American journal of respiratory and critical care medicine. 2007 Sep 15;176(6):532–55.

24. Global Initiative for Chronic Obstructive Lung Disease (Updated 2014). 2014; Available from: http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html

25. Agusti A, Edwards LD, Celli B, Macnee W, Calverley PM, Mullerova H, et al. Characteristics, stability and outcomes of the 2011 GOLD COPD groups in the ECLIPSE cohort. The European respiratory journal. 2013 Sep;42(3):636–46.

26. Agusti A, Hurd S, Jones P, Fabbri LM, Martinez F, Vogelmeier C, et al. FAQs about the GOLD 2011 assessment proposal of COPD: a comparative analysis of four different cohorts. The European respiratory journal. 2013 Nov;42(5):1391–401.

27. Leivseth L, Brumpton BM, Nilsen TI, Mai XM, Johnsen R, Langhammer A. GOLD classifications and mortality in chronic obstructive pulmonary disease: the HUNT Study, Norway. Thorax. 2013 Oct;68(10):914–21.

28. Govaerts E, Demedts M, Van de Woestijne KP. Total respiratory impedance and early emphysema. The European respiratory journal. 1993 Sep;6(8):1181–5.

29. Gould GA, Redpath AT, Ryan M, Warren PM, Best JJ, Cameron EJ, et al. Parenchymal emphysema measured by CT lung density correlates with lung function in patients with bullous disease. The European respiratory journal. 1993 May;6(5):698–704.

30. Hersh CP, Washko GR, Estepar RS, Lutz S, Friedman PJ, Han MK, et al. Paired inspiratory-expiratory chest CT scans to assess for small airways disease in COPD. Respiratory research. 2013;14:42.

31. Smith BM, Hoffman EA, Basner RC, Kawut SM, Kalhan R, Barr RG. Not all Measures of Hyperinflation are Created Equal: Lung Structure and Clinical Correlates of Gas-Trapping and Hyper-Expansion in COPD: The Multi-Ethnic Study of Atherosclerosis Study (MESA) COPD Study. Chest [Internet]. 2014 Jan 30; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24481056

32. Kim V, Davey A, Comellas AP, Han MK, Washko G, Martinez CH, et al. Clinical and computed tomographic predictors of chronic bronchitis in COPD: a cross sectional analysis of the COPDGene study. Respiratory research. 2014 Apr 27;15(1):52.

33. Park HY, Churg A, Wright JL, Li Y, Tam S, Man SFP, et al. Club cell protein 16 and disease progression in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013 Dec 15;188(12):1413–9.

34. Kim V, Criner GJ. Chronic bronchitis and chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2013 Feb 1;187(3):228–37.

35. Teculescu D. [Clear definitions of chronic bronchopulmonary diseases: a necessity for practice, teaching and research in pneumology]. Revue de pneumologie clinique. 1990;46(5):194–9.

36. Thurlbeck WM, Angus GE. A Distribution Curve for Chronic Bronchitis. Thorax. 1964 Sep;19:436–42.

37. Fletcher CM, Peto R, Tinker CM. A comparison of the assessment of simple bronchitis (chronic mucus hypersecretion) by measurements of sputum volume and by standardized questions on phlegm production. International journal of epidemiology. 1974 Dec;3(4):315–9.

38. Lumsden AB, McLean A, Lamb D. Goblet and Clara cells of human distal airways: evidence for smoking induced changes in their numbers. Thorax. 1984 Nov;39(11):844–9.

39. Yoshida T, Tuder RM. Pathobiology of cigarette smoke-induced chronic obstructive pulmonary disease. Physiological reviews. 2007 Jul;87(3):1047–82.

40. Reid L. Measurement of the bronchial mucous gland layer: a diagnostic yardstick in chronic bronchitis. Thorax. 1960 Jun;15:132–41.

41. Cosio MG, Cosio Piqueras MG. Pathology of emphysema in chronic obstructive pulmonary disease. Monaldi archives for chest disease = Archivio Monaldi per le malattie del torace / Fondazione clinica del lavoro, IRCCS [and] Istituto di clinica tisiologica e malattie apparato respiratorio, Universita di Napoli, Secondo ateneo. 2000 Apr;55(2):124–9.

42. Takahashi M, Fukuoka J, Nitta N, Takazakura R, Nagatani Y, Murakami Y, et al. Imaging of pulmonary emphysema: a pictorial review. International journal of chronic obstructive pulmonary disease. 2008;3(2):193–204.

43. Ryder RC, Dunnill MS, Anderson JA. A quantitative study of bronchial mucous gland volume, emphysema and smoking in a necropsy population. The Journal of pathology. 1971 May;104(1):59–71.

44. Anderson JA, Dunnill MS, Ryder RC. Dependence of the incidence of emphysema on smoking history, age, and sex. Thorax. 1972 Sep;27(5):547–51.

45. Wright JL, Barry W, Pare PD, Hogg JC. Ranking the severity of emphysema on whole lung slices. Concordance of upper lobe, lower lobe, and entire lung ranks. The American review of respiratory disease. 1986 May;133(5):930–1.

46. Sanders C, Nath PH, Bailey WC. Detection of emphysema with computed tomography. Correlation with pulmonary function tests and chest radiography. Investigative radiology. 1988 Apr;23(4):262–6.

47. Petty TL, Silvers GW, Stanford RE. Mild emphysema is associated with reduced elastic recoil and increased lung size but not with air-flow limitation. The American review of respiratory disease. 1987 Oct;136(4):867–71.

48. Leopold JG, Gough J. The centrilobular form of hypertrophic emphysema and its relation to chronic bronchitis. Thorax. 1957 Sep;12(3):219–35.

49. Thurlbeck WM. The incidence of pulmonary emphysema, with observations on the relative incidence and spatial distribution of various types of emphysema. The American review of respiratory disease. 1963 Feb;87:206–15.

50. Thurlbeck WM, Churg AM. Pathology of the lung. 2nd ed. Thieme Medical Publishers; 1995.

51. Peters RM, Peters BA, Benirschke SK, Friedman PJ. Chest dimensions in young adults with spontaneous pneumothorax. The Annals of thoracic surgery. 1978 Mar;25(3):193–6.

52. Bates, D.V., Macklem, P.T., Christie, R.V. Respiratory function in disease. Second. W.B. Saunders Company; 1971. 584 p.

53. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. The New England journal of medicine. 2004 Jun 24;350(26):2645–53.

54. Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet. 2004 Aug 21;364(9435):709–21.

55. Hogg JC, McDonough JE, Gosselink JV, Hayashi S. What drives the peripheral lung-remodeling process in chronic obstructive pulmonary disease? Proceedings of the American Thoracic Society. 2009 Dec;6(8):668–72.

56. Hogg JC, McDonough JE, Sanchez PG, Cooper JD, Coxson HO, Elliott WM, et al. Micro-computed tomography measurements of peripheral lung pathology in chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society. 2009 Sep 15;6(6):546–9.

57. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. The New England journal of medicine. 2011 Oct 27;365(17):1567–75.

58. Hogg JC. A brief review of chronic obstructive pulmonary disease. Canadian respiratory journal : journal of the Canadian Thoracic Society. 2012 Nov;19(6):381–4.

59. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. Nat Med. 2012 Nov;18(11):1711–5.

60. Caramori G, Adcock IM, Di Stefano A, Chung KF. Cytokine inhibition in the treatment of COPD. International journal of chronic obstructive pulmonary disease. 2014;9:397–412.

61. Paats MS, Bergen IM, Hoogsteden HC, van der Eerden MM, Hendriks RW. Systemic CD4+ and CD8+ T-cell cytokine profiles correlate with GOLD stage in stable COPD. The European respiratory journal. 2012 Aug;40(2):330–7.

62. Barnes PJ. Mediators of chronic obstructive pulmonary disease. Pharmacological reviews. 2004 Dec;56(4):515–48.

63. Visperas A, Do J, Min B. Cellular Factors Targeting APCs to Modulate Adaptive T Cell Immunity. Journal of immunology research. 2014;2014:750374.

64. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Annals of internal medicine. 1987 Feb;106(2):196–204.

65. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2000 May;161(5):1608–13.

66. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. Chest. 2000 May;117(5 Suppl 2):398S – 401S.

67. Goldstein R, Brooks D. Pulmonary rehabilitation at the time of the COPD exacerbation. Clinics in chest medicine. 2014 Jun;35(2):391–8.

68. Wilson R. Bacteria and airway inflammation in chronic obstructive pulmonary disease: more evidence. American journal of respiratory and critical care medicine. 2005 Jul 15;172(2):147–8.

69. Wilson R, Sethi S, Anzueto A, Miravitlles M. Antibiotics for treatment and prevention of exacerbations of chronic obstructive pulmonary disease. The Journal of infection. 2013 Dec;67(6):497–515.

70. Mackay AJ, Hurst JR. COPD exacerbations: causes, prevention, and treatment. Immunology and allergy clinics of North America. 2013 Feb;33(1):95–115.

71. Anderson HR, Spix C, Medina S, Schouten JP, Castellsague J, Rossi G, et al. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project. The European respiratory journal. 1997 May;10(5):1064–71.

72. Sethi S. Infectious etiology of acute exacerbations of chronic bronchitis. Chest. 2000 May;117(5 Suppl 2):380S – 5S.

73. Agusti A, Sin DD. Biomarkers in COPD. Clinics in chest medicine. 2014 Mar;35(1):131–41.

74. Baraldo S, Turato G, Lunardi F, Bazzan E, Schiavon M, Ferrarotti I, et al. Immune activation in α1-antitrypsin-deficiency emphysema. Beyond the protease-antiprotease paradigm. Am J Respir Crit Care Med. 2015 Feb 15;191(4):402–9.

75. Stockley RA. Alpha1-antitrypsin review. Clinics in chest medicine. 2014 Mar;35(1):39–50.

76. Demkow U, van Overveld FJ. Role of elastases in the pathogenesis of chronic obstructive pulmonary disease: implications for treatment. European journal of medical research. 2010 Nov 4;15 Suppl 2:27–35.

77. Hurst JR, Vestbo J, Anzueto A, Locantore N, Mullerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. The New England journal of medicine. 2010 Sep 16;363(12):1128–38.

78. Sethi S, Murphy TF. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. The New England journal of medicine. 2008 Nov 27;359(22):2355–65.

79. Tan WC, Bourbeau J, Hernandez P, Chapman KR, Cowie R, FitzGerald JM, et al. Exacerbation-like respiratory symptoms in individuals without chronic obstructive pulmonary disease: results from a population-based study. Thorax. 2014 Aug;69(8):709–17.

80. Silva GE, Sherrill DL, Guerra S, Barbee RA. Asthma as a risk factor for COPD in a longitudinal study. Chest. 2004 Jul;126(1):59–65.

81. Barnes PJ. Against the Dutch hypothesis: asthma and chronic obstructive pulmonary disease are distinct diseases. American journal of respiratory and critical care medicine. 2006 Aug 1;174(3):240–3; discussion 243–4.

82. Barnes PJ. Cellular and molecular mechanisms of chronic obstructive pulmonary disease. Clinics in chest medicine. 2014 Mar;35(1):71–86.

83. James AL, Palmer LJ, Kicic E, Maxwell PS, Lagan SE, Ryan GF, et al. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. American journal of respiratory and critical care medicine. 2005 Jan 15;171(2):109–14.

84. Thomson NC, Chaudhuri R, Livingston E. Asthma and cigarette smoking. The European respiratory journal. 2004 Nov;24(5):822–33.

85. Torsvik V, Ovreas L. Microbial diversity and function in soil: from genes to ecosystems. Current opinion in microbiology. 2002 Jun;5(3):240–5.

86. Relman DA. New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. The Journal of infectious diseases. 2002 Dec 1;186 Suppl 2:S254–8.

87. Shanahan F. The host-microbe interface within the gut. Best practice & research Clinical gastroenterology. 2002 Dec;16(6):915–31.

88. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, et al. Viral and microbial community dynamics in four aquatic environments. The ISME journal. 2010 Jun;4(6):739–51.

89. Dollive S, Chen YY, Grunberg S, Bittinger K, Hoffmann C, Vandivier L, et al. Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment. PloS one. 2013;8(8):e71806.

90. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007 Oct 18;449(7164):804–10.

91. Li K, Bihan M, Yooseph S, Methe BA. Analyses of the microbial diversity across the human microbiome. PloS one. 2012;7(6):e32118.

92. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PloS one. 2013;8(6):e66019.

93. Scupham AJ, Presley LL, Wei B, Bent E, Griffith N, McPherson M, et al. Abundant and diverse fungal microbiota in the murine intestine. Applied and environmental microbiology. 2006 Jan;72(1):793–801.

94. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010 Jul 15;466(7304):334–8.

95. De Dombal FT, Burch PR, Watkinson G. Aetiology of ulcerative colitis. Gut. 1969 Apr;10(4):270–7.

96. Macfarlane GT, Macfarlane S. Human colonic microbiota: ecology, physiology and metabolic potential of intestinal bacteria. Scandinavian journal of gastroenterology Supplement. 1997;222:3–9.

97. Nesbitt B. On the Presence of Cholin and Neurin in the Intestinal Canal during Its Complete Obstruction : A Research on Autointoxication. The Journal of experimental medicine. 1899 Jan 1;4(1):1–18.

98. Khoury KA, Floch MH, Herskovic T. Effects of neomycin and penicillin administration on mucosal proliferation of the mouse small intestine. With morphological and functional correlations. The Journal of experimental medicine. 1969 May 1;129(5):1063–78.

99. Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS, et al. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proceedings of the National Academy of Sciences of the United States of America. 2011 Mar 29;108(13):5354–9.

100. Ferreira RB, Gill N, Willing BP, Antunes LC, Russell SL, Croxen MA, et al. The intestinal microbiota plays a role in Salmonella-induced colitis independent of pathogen colonization. PloS one. 2011;6(5):e20338.

101. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the “healthy” smoker and in COPD. PloS one. 2011;6(2):e16384.

102. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PLoS One. 2012;7(10):1–10.

103. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. American journal of respiratory and critical care medicine. 2013 May 15;187(10):1067–75.

104. Bartlett JG, Chang TW, Gurwith M, Gorbach SL, Onderdonk AB. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. The New England journal of medicine. 1978 Mar 9;298(10):531–4.

105. Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. Gut. 2013 Oct;62(10):1505–10.

106. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013 Aug 29;500(7464):541–6.

107. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. PloS one. 2010;5(1):e8578.

108. Huang YJ, Kim E, Cox MJ, Brodie EL, Brown R, Wiener-Kronish JP, et al. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. Omics : a journal of integrative biology. 2010 Feb;14(1):9–59.

109. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2012 May 15;185(10):1073–80.

110. Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Kehagia V, Jones GR, et al. Bacterial activity in cystic fibrosis lung infections. Respiratory research. 2005;6:49.

111. Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, et al. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. Thorax. 2013 Aug;68(8):731–7.

112. Pezzulo AA, Kelly PH, Nassar BS, Rutland CJ, Gansemer ND, Dohrn CL, et al. Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. Applied and environmental microbiology. 2013 Oct;79(19):5936–41.

113. Rogers GB, Marsh P, Stressmann AF, Allen CE, Daniels TV, Carroll MP, et al. The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2010 Nov;16(11):1656–8.

114. Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diagnostic microbiology and infectious disease. 2008 Oct;62(2):133–41.

115. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1:1–12.

116. Satir P, Sleigh MA. The physiology of cilia and mucociliary interactions. Annual review of physiology. 1990;52:137–55.

117. Smith CM, Kulkarni H, Radhakrishnan P, Rutman A, Bankart MJ, Williams G, et al. Ciliary dyskinesia is an early feature of respiratory syncytial virus infection. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology [Internet]. 2013 Mar 21; Available from: http://www.ncbi.nlm.nih.gov/pubmed/23520320

118. Sagel SD, Davis SD, Campisi P, Dell SD. Update of respiratory tract disease in children with primary ciliary dyskinesia. Proceedings of the American Thoracic Society. 2011 Sep;8(5):438–43.

119. Tambascio J, Lisboa RM, Passarelli Rde C, Martinez JA, Gastaldi AC. Adhesiveness and purulence of respiratory secretions: implications for mucociliary transport in patients with bronchiectasis. Jornal

brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia. 2010 Sep;36(5):545–53.

120. Sibille Y, Reynolds HY. Macrophages and polymorphonuclear neutrophils in lung defense and injury. The American review of respiratory disease. 1990 Feb;141(2):471–501.

121. Hogg JC, van Eeden S. Pulmonary and systemic response to atmospheric pollution. Respirology. 2009 Apr;14(3):336–46.

122. Gleeson K, Eggli DF, Maxwell SL. Quantitative aspiration during sleep in normal subjects. Chest. 1997 May;111(5):1266–72.

123. Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, et al. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. Proceedings of the National Academy of Sciences of the United States of America. 2012 Aug 21;109(34):13769–74.

124. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. American journal of respiratory and critical care medicine. 2011 Oct 15;184(8):957–63.

125. Ramakrishnan VR, Ferril GR, Suh JD, Woodson T, Green TJ, Kingdom TT. Upper and lower airways associations in patients with chronic rhinosinusitis and bronchiectasis. International forum of allergy & rhinology. 2013 Nov;3(11):921–7.

126. Berkhout MC, Rijntjes E, El Bouazzaoui LH, Fokkens WJ, Brimicombe RW, Heijerman HG. Importance of bacteriology in upper airways of patients with Cystic Fibrosis. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. 2013 Sep;12(5):525–9.

127. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. PloS one. 2010;5(12):e15216.

128. Lourenco RV, Klimek MF, Borowski CJ. Deposition and clearance of 2 micron particles in the tracheobronchial tree of normal subjects--smokers and nonsmokers. The Journal of clinical investigation. 1971 Jul;50(7):1411–20.

129. Smaldone GC, Foster WM, O'Riordan TG, Messina MS, Perry RJ, Langenback EG. Regional impairment of mucociliary clearance in chronic obstructive pulmonary disease. Chest. 1993 May;103(5):1390–6.

130. Fanta CH. Clinical aspects of mucus and mucous plugging in asthma. The Journal of asthma : official journal of the Association for the Care of Asthma. 1985;22(6):295–301.

131. Leopold PL, O'Mahony MJ, Lian XJ, Tilley AE, Harvey BG, Crystal RG. Smoking is associated with shortened airway cilia. PloS one. 2009;4(12):e8157.

132. Larsen JM, Steen-Jensen DB, Laursen JM, Sondergaard JN, Musavian HS, Butt TM, et al. Divergent pro-inflammatory profile of human dendritic cells in response to commensal and pathogenic bacteria associated with the airway microbiota. PloS one. 2012;7(2):e31976.

133. Forsythe P, Inman MD, Bienenstock J. Oral treatment with live Lactobacillus reuteri inhibits the allergic airway response in mice. American journal of respiratory and critical care medicine. 2007 Mar 15;175(6):561–9.

134. Karimi K, Inman MD, Bienenstock J, Forsythe P. Lactobacillus reuteri-induced regulatory T cells protect against an allergic airway response in mice. American journal of respiratory and critical care medicine. 2009 Feb 1;179(3):186–93.

135. Eishi Y. Etiologic aspect of sarcoidosis as an allergic endogenous infection caused by Propionibacterium acnes. BioMed research international. 2013;2013:935289.

136. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, et al. Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. Ann Am Thorac Soc. 2015 Mar 24;

137. Di Stefano A, Turato G, Maestrelli P, Mapp CE, Ruggieri MP, Roggeri A, et al. Airflow limitation in chronic bronchitis is associated with T-lymphocyte and macrophage infiltration of the bronchial mucosa. American journal of respiratory and critical care medicine. 1996 Feb;153(2):629–32.

138. Suzuki M, Elliott WM, John EM, Shizu H, Pablo GS, Joel DC, et al. The Quantitative Relationship Between Alveolar Inflammation And Emphysematous Destruction In COPD. In: B105 CHRONIC OBSTRUCTIVE PULMONARY DISEASE: PATHOGENESIS AND MECHANISMS OF DISEASE [Internet]. American Thoracic Society; [cited 2013 Oct 15]. p. A3872–A3872. Available from: http://dx.doi.org/10.1164/ajrccm-conference.2010.181.1_MeetingAbstracts.A3872

139. Suzuki M, Elliott WM, John EM, Shizu H, Pablo GS, Joel DC, et al. The Quantitative Relationship Between Collagen Deposition And Lung Structure In COPD. In: B37 CHRONIC OBSTRUCTIVE PULMONARY DISEASE PATHOGENESIS I [Internet]. American Thoracic Society; [cited 2013 Oct 15]. p. A6808–A6808. Available from: http://dx.doi.org/10.1164/ajrccm-conference.2010.181.1_MeetingAbstracts.A6808

140. Kristanto W, van Ooijen PM, Groen JM, Vliegenthart R, Oudkerk M. Small calcified coronary atherosclerotic plaque simulation model: minimal size and attenuation detectable by 64-MDCT and MicroCT. The international journal of cardiovascular imaging. 2012 Apr;28(4):843–53.

141. Low TB, Greene CM, O'Neill SJ, McElvaney NG. Quantification and evaluation of the role of antielastin autoantibodies in the emphysematous lung. Pulmonary medicine. 2011;2011:826160.

142. Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. The New England journal of medicine. 2002 Aug 15;347(7):465–71.

143. De Serres G, Lampron N, La Forge J, Rouleau I, Bourbeau J, Weiss K, et al. Importance of viral and bacterial infections in chronic obstructive pulmonary disease exacerbations. Journal of clinical

virology : the official publication of the Pan American Society for Clinical Virology. 2009 Oct;46(2):129–33.

144. Utokaparch S, Sze MA, Gosselink JV, McDonough JE, Elliott WM, Hogg JC, et al. Respiratory viral detection and small airway inflammation in lung tissue of patients with stable, mild COPD. Copd. 2014 Apr;11(2):197–203.

145. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio. 2015;6(2):e00037.

146. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. Lancet. 2014 Aug 23;384(9944):691–702.

147. Whittaker RH. Evolution and measurement of species diversity. Taxon. 1972 May;21(2/3):213–51.

148. McArdle BH, Anderson MJ. FITTING MULTIVARIATE MODELS TO COMMUNITY DATA: A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. Ecology. 2001;82(Journal Article):290–7.

149. Searle SR, Casella G, McCulloch CE. Variance Components. Wiley-Interscience; 1992.

150. Oksanen, Jari, Blanchet, Guillaume F, Kindt, Roeland, Legendre, Pierre, Minchin, Peter R, O'Hara, RB, et al. Community Ecology Package [Internet]. 2015. Available from: http://cran.r-project.org, https://github.com/vegandevs/vegan

151. Breiman L. Random Forests. Machine Learning. 2001 Oct 1;45(1):5–32.

152. Kursa MB, Rudnicki WR. Feature selection with the boruta package [Internet]. Journal; 2010 [cited 2013 Feb 9]. Available from: http://www.jstatsoft.org/v36/i11/paper/

153. S D, Gonzalez I, K-A LC, Monget P, Coquery J, Yao F, et al. mixOmics: Omics data integration project. R Package version 41-4 [Internet]. 2013; Available from: http://CRAN.R-project.org/package=mixOmics

154. Saetta M, Baraldo S, Corbino L, Turato G, Braccioni F, Rea F, et al. CD8+ve cells in the lungs of smokers with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999 Aug;160(2):711–7.

155. Siena L, Gjomarkaj M, Elliot J, Pace E, Bruno A, Baraldo S, et al. Reduced apoptosis of CD8+ T-lymphocytes in the airways of smokers with mild/moderate COPD. Respir Med. 2011 Oct;105(10):1491–500.

156. Huber M, Lohoff M. Change of paradigm: CD8+ T cells as important helper for CD4+ T cells during asthma and autoimmune encephalomyelitis. Allergo J Int. 2015;24(1):8–15.

157. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.

158. Schneider JP, Ochs M. Alterations of mouse lung tissue dimensions during processing for morphometry: a comparison of methods. American journal of physiology Lung cellular and molecular physiology. 2014 Feb 15;306(4):L341–50.

159. Kumar K, Manohar MR. Consistency of histomorphometric measurements of normal oral epithelium subjected to variations in routine tissue processing. Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology. 2011 Feb;33(1):12–8.

160. Hsu PK, Huang HC, Hsieh CC, Hsu HS, Wu YC, Huang MH, et al. Effect of formalin fixation on tumor size determination in stage I non-small cell lung cancer. The Annals of thoracic surgery. 2007 Dec;84(6):1825–9.

161. Saetta M, Di Stefano A, Maestrelli P, Ferraresso A, Drigo R, Potena A, et al. Activated T-lymphocytes and macrophages in bronchial mucosa of subjects with chronic bronchitis. The American review of respiratory disease. 1993 Feb;147(2):301–6.

162. Baraldo S, Turato G, Saetta M. Pathophysiology of the small airways in chronic obstructive pulmonary disease. Respiration; international review of thoracic diseases. 2012;84(2):89–97.

163. Fletcher CM. Chronic Bronchitis - Its Prevalence, Nature, and Pathogenesis. Am Rev Respir Dis. 1959;80(4):483–94.

164. Fletcher C, Peto R. The natural history of chronic airflow obstruction. British medical journal. 1977 Jun 25;1(6077):1645–8.

165. Li K, Bihan M, Methe BA. Analyses of the stability and core taxonomic memberships of the human microbiome. PloS one. 2013;8(5):e63139.

166. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS biology. 2012;10(8):e1001377.

167. Conlan S, Kong HH, Segre JA. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. PloS one. 2012;7(10):e47075.

168. Morgan XC, Segata N, Huttenhower C. Biodiversity and functional genomics in the human microbiome. Trends in genetics : TIG. 2013 Jan;29(1):51–8.

169. Sze, M, Dimitriu P.A., Suzuki M, McDonough, J.E., Campbell, J.D., Brothers, J.F., et al. The host response to the bacterial lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;189(A1012).

170. Sze, M, Erb-Downward, J.R., Hayashi, S, Elliott, W.M., Sin, D.D., Huffnagle, G.B., et al. Is there a difference? A comparison between a nested and touchdown PCR approach for bacterial microbiome analysis. Am J Respir Crit Care Med. 2014;189(A4969).

171. Sze, M, Dimitriu P.A., McDonough, J.E., Suzuki, M, Gosselink, J.V., Elliott, W.M., et al. The host response to the lung microbiome during emphysematous destruction. Am J Respir Crit Care Med. 2013;187(A3493).

172. Sze, M, Dimitriu P.A., McDonough, J.E., Suzuki, M, Gosselink, J.V., Elliott, W.M., et al. The host response to the lung microbiome in lung tissue undergoing emphysematous destruction. Ann Am Thorac Soc. 2013 Jan;Suppl 1(S77).

173. Gosselink JV, Hayashi S, Elliott WM, Xing L, Chan B, Yang L, et al. Differential expression of tissue repair genes in the pathogenesis of chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2010 Jun 15;181(12):1329–35.

174. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. PLoS ONE. 2012;7(6):e39315.

175. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one. 2011;6(12):e27310.

176. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature protocols. 2008;3(9):1452–6.

177. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, et al. Changes in the lung microbiome following lung transplantation include the emergence of two distinct Pseudomonas species with distinct clinical associations. PLoS ONE. 2014;9(5):e97214.

178. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology. 2003;4(5):P3.

179. Dixon P. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 2003 Dec;14(6):927–30.

180. Lysenko ES, Ratner AJ, Nelson AL, Weiser JN. The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. PLoS Pathog. 2005 Sep;1(1):e1.

181. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clinical microbiology reviews. 1996 Jan;9(1):18–33.

182. Nakou A, Papaparaskevas J, Diamantea F, Skarmoutsou N, Polychronopoulos V, Tsakris A. A prospective study on bacterial and atypical etiology of acute exacerbation in chronic obstructive pulmonary disease. Future Microbiol. 2014;9(11):1251–60.

183. Aydemir Y, Aydemir Ö, Kalem F. Relationship between the GOLD combined COPD assessment staging system and bacterial isolation. Int J Chron Obstruct Pulmon Dis. 2014;9:1045–51.

184. Zhang X, Liu G, Lenburg ME, Spira A. Comparison of smoking-induced gene expression on Affymetrix Exon and 3'-based expression arrays. Genome informatics International Conference on Genome Informatics. 2007;18:247–57.

185. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J. Metagenomic pyrosequencing and microbial identification. Clinical chemistry. 2009 May;55(5):856–66.

186. Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, et al. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic acids research. 2007;35(19):e130.

187. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, et al. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Meth. 2009;6(9):639–41.

188. Hartmann M, Howes CG, Abarenkov K, Mohn WW, Nilsson RH. V-Xtractor: An open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16 S/18 S) ribosomal RNA gene sequences. Journal of microbiological methods. 2010;83(2):250–3.

189. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194–200.

190. Hartmann M, Howes CG, VanInsberghe D, Yu H, Bachar D, Christen R, et al. Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. The ISME Journal [Internet]. 2012; Available from: http://www.nature.com/ismej/journal/vaop/ncurrent/full/ismej201284a.html

191. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. ApplEnvironMicrobiol. 2007;73(16):5261–7.

192. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012 Mar;6(3):610–8.

193. Werner JJ, Koren O, Hugenholtz P, Desantis TZ, Walters WA, Caporaso JG, et al. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME Journal [Internet]. 2011; Available from: http://www.ncbi.nlm.nih.gov/pubmed/21716311

194. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. ApplEnvironMicrobiol. 2009;75(23):7537–41.

195. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80.

196. Riehle K, Coarfa C, Jackson A, Ma J, Tandon A, Paithankar S, et al. The Genboree Microbiome Toolset and the analysis of 16S rRNA microbial sequences. BMC Bioinformatics [Internet]. 2012 Aug 24 [cited 2012 Oct 12];13(Suppl 13). Available from: http://www.biomedcentral.com/1471-2105/13/S13/S11/abstract

197. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry. 2009 Apr;55(4):611–22.

198. Brankatschk R, Bodenhausen N, Zeyer J, Burgmann H. Simple absolute quantification method correcting for quantitative PCR efficiency variations for microbial community samples. Applied and environmental microbiology. 2012 Jun;78(12):4481–9.

199. Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. Journal of biomolecular techniques : JBT. 2004 Sep;15(3):155–66.

200. Suzuki MT, Giovannoni SJ. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and environmental microbiology. 1996 Feb;62(2):625–30.

201. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical chemistry. 2011 Nov 15;83(22):8604–10.

202. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Analytical chemistry. 2012 Jan 17;84(2):1003–11.

203. Strain MC, Richman DD. New assays for monitoring residual HIV burden in effectively treated individuals. Current opinion in HIV and AIDS. 2013 Mar;8(2):106–10.

204. Laboratories B-R. Droplet digitalTM PCR Applications Guide. :94–5.

205. Ma J, Li N, Guarnera M, Jiang F. Quantification of Plasma miRNAs by Digital PCR for Cancer Diagnosis. Biomarker insights. 2013;8:127–36.

206. Roberts CH, Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M, et al. Development and evaluation of a next-generation digital PCR diagnostic assay for ocular Chlamydia trachomatis infections. Journal of clinical microbiology. 2013 Jul;51(7):2195–203.

207. Han MK, Huang YJ, Lipuma JJ, Boushey HA, Boucher RC, Cookson WO, et al. Significance of the microbiome in obstructive lung disease. Thorax. 2012 May;67(5):456–63.

208. Kiley JP, Caler EV. The lung microbiome. A new frontier in pulmonary medicine. Annals of the American Thoracic Society. 2014 Jan;11 Suppl 1:S66–70.

209. Sze MA, Hogg JC, Sin DD. Bacterial microbiome of lungs in COPD. International journal of chronic obstructive pulmonary disease. 2014;9:229–38.

210. Sethi S. Bacterial infection and the pathogenesis of COPD. Chest. 2000 May;117(5 Suppl 1):286S – 91S.

211. Ott SJ, Musfeldt M, Ullmann U, Hampe J, Schreiber S. Quantification of intestinal bacterial populations by real-time PCR with a universal primer set and minor groove binder probes: a global approach to the enteric flora. Journal of clinical microbiology. 2004 Jun;42(6):2566–72.

212. Tajadini M, Panjehpour M, Javanmard SH. Comparison of SYBR Green and TaqMan methods in quantitative real-time polymerase chain reaction analysis of four adenosine receptor subtypes. Advanced biomedical research. 2014;3:85.

213. Hogg JC, Macklem PT, Thurlbeck WM. Site and nature of airway obstruction in chronic obstructive lung disease. The New England journal of medicine. 1968 Jun 20;278(25):1355–60.

214. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. The Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2015 May 6;

215. Herr C, Han G, Li D, Tschernig T, Dinh QT, Beißwenger C, et al. Combined exposure to bacteria and cigarette smoke resembles characteristic phenotypes of human COPD in a murine disease model. Exp Toxicol Pathol. 2015 Jan 15;

216. Finney LJ, Ritchie A, Pollard E, Johnston SL, Mallia P. Lower airway colonization and inflammatory response in COPD: a focus on Haemophilus influenzae. Int J Chron Obstruct Pulmon Dis. 2014;9:1119–32.

217. Wang X, Mair R, Hatcher C, Theodore MJ, Edmond K, Wu HM, et al. Detection of bacterial pathogens in Mongolia meningitis surveillance with a new real-time PCR assay to detect Haemophilus influenzae. Int J Med Microbiol. 2011 Apr;301(4):303–9.

218. Sze MA, Abbasi M, Hogg JC, Sin DD. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD GOLD 2. PLoS ONE. 2014;9(10):e110351.

219. King PT, Lim S, Pick A, Ngui J, Prodanovic Z, Downey W, et al. Lung T-cell responses to nontypeable Haemophilus influenzae in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2013 May;131(5):1314–21.e14.

220. Vestbo J, Prescott E, Lange P. Association of chronic mucus hypersecretion with FEV1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. American journal of respiratory and critical care medicine. 1996 May;153(5):1530–5.

221. Lugade AA, Bogner PN, Thatcher TH, Sime PJ, Phipps RP, Thanavala Y. Cigarette smoke exposure exacerbates lung inflammation and compromises immunity to bacterial infection. J Immunol. 2014 Jun 1;192(11):5226–35.

222. Teo E, House H, Lockhart K, Purchuri SN, Pushparajah J, Cripps AW, et al. Haemophilus influenzae oral vaccination for preventing acute exacerbations of chronic bronchitis and chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2014;9:CD010010.

223. Molyneaux PL, Mallia P, Cox MJ, Footitt J, Willis-Owen SAG, Homola D, et al. Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013 Nov 15;188(10):1224–31.

224. Desai H, Eschberger K, Wrona C, Grove L, Agrawal A, Grant B, et al. Bacterial colonization increases daily symptoms in patients with chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2014 Mar;11(3):303–9.

225. Geelen TH, Gaajetaan GR, Wouters EF, Rohde GG, Franssen FM, Grauls GE, et al. The host immune response contributes to Haemophilus influenzae virulence. Respir Med. 2014 Jan;108(1):144–52.

226. Millares L, Marin A, Garcia-Aymerich J, Sauleda J, Belda J, Monsó E, et al. Specific IgA and metalloproteinase activity in bronchial secretions from stable chronic obstructive pulmonary disease patients colonized by Haemophilus influenzae. Respir Res. 2012;13:113.

227. Zenobia C, Hajishengallis G. Basic biology and role of interleukin-17 in immunity and inflammation. Periodontol 2000. 2015 Oct;69(1):142–59.

228. Pereira JP, Kelly LM, Cyster JG. Finding the right niche: B-cell migration in the early phases of T-dependent antibody responses. Int Immunol. 2010 Jun;22(6):413–9.

Data Appendix 1 List of significant genes correlated with Specific Phyla Using the Michigan Method (Protocol 1). Table relates to Chapter 4 results.

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
CD24	GAS7
FMO3	IGF1
PLAC8	NLRP3
TRIM32	MT1IP
ZFP90	METTL7B
SNRPN	MS4A6E
ZNF211	ABCC2
ECHDC2	ZNF179
SFTPC	TRIM7
EPB49	SLC22A16
RCAN3	RAC3
NPAL3	CD300E
CHL1	OLFML2B
TUSC3	HPD
CYP4B1	PI3
TFDP2	GPR173
MGC50559	SNX18
PKHD1L1	CES7
ZNF417	CACNA1E
MAPK10	SC65
IGBP1	C16orf82
C7	ADAMTSL4
FOXO4	TFF1
ZNF138	APRT
ZSCAN18	KIAA0774
NXF3	C10orf90
FOLR1	LDLRAD3
FLJ45803	NR5A1
FGFR2	CEACAM3
CPAMD8	NAB1
PGM5	DDR1
ING4	KRT8
PERP	MTMR1
ULK2	PLAGL2
CHD6	PRSS12
CTSO	STAM2

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
EPC1	TMTC2
IQCA	SPINT2
ZNF84	NCALD
TRPC4	EFHC1
SEPT4	DLG5
RAD52	ARL6IP6
C8orf70	SLFN13
PLEKHA5	C10orf57
ITGA10	MID1
NCALD	ABCA5
TAS2R5	GPR177
KLK11	DOPEY1
OSBPL6	BMP3
LASS4	ECHDC2
SLFN13	DNAJC6
ANTXR1	UXS1
CRIP1	SERTAD4
C5orf42	NCOA7
PRELP	RNF2
CYP4X1	SCRN1
TPPP3	BBS5
PXDNL	TDRD10
CYBRD1	FARP1
CTGF	MUC15
C6orf204	UST
MAP3K13	CLIC5
RP1-21O18.1	TP63
CSRP1	ZNF417
CTDSPL	ZER1
CELSR1	KLF5
PPM1K	C11orf49
GCC2	RASSF9
SLC25A29	SATB1
SOX13	BAIAP2L1
NLRX1	CAPRIN2
TRIM68	ANK3
GEN1	OCIAD2
PUS10	C18orf10
ISYNA1	PLEKHA5

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
CX3CL1	PARD6B
KBTBD3	ACSS1
HNT	C9orf45
SPON1	ZNF345
RIPK5	ERV3
C2orf40	CHD6
ZC3H6	CLSTN1
ZBTB33	PIGQ
TIMP4	CEP97
WEE1	SPEF2
KIAA1704	TUFT1
GOLPH3L	C19orf33
ZNF483	ZNF295
TEAD1	RFXDC2
INMT	C16orf80
LOC492311	BTBD6
EXPH5	POGZ
FST	transcript 3184925, GenBank AK054857
SMARCA2	HIVEP1
BTBD6	C14orf159
RTTN	C5orf42
PRELID2	ILDR1
BBS1	PERP
ST6GALNAC2	FAM107B
MAOB	C11orf63
FLNB	ZNF678
CGNL1	PPP2R3A
MUC15	C8orf70
GALT	FUZ
PRDM11	C9orf93
C4orf34	KIAA1407
KIAA1841	PER3
EPHX2	C4orf31
SCRN1	PCTK3
TMEM159	USP28
FOXP1	FANCL
GPR126	LIMA1
PSD3	TFDP2
ST6GALNAC6	FMO3

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
EFHA2	AGBL5
RNF125	THEM4
TTC12	ZNF713
C11orf52	NPAT
PLEKHB1	FOXP1
TDRD10	GOLPH3L
STRBP	CD24
CLIC5	IDS
ZNF132	AKAP9
ZNF461	GNRH1
A2M	TUSC3
MUC20	KIAA1324
PTPN13	CDH1
tcag7.1177	SCNN1G
ZNF678	OSBPL6
FCGBP	ST3GAL4
FHL1	C11orf60
C6orf155	JARID1B
KIAA1407	AQP3
SRPX2	IFT52
AEBP2	DNHD1
AKAP9	C11orf52
RNF2	ZDHHC13
CLIC5	TTC12
MGC24039	CHIA
MYO6	LRBA
GPR177	MAP3K13
PGDS	MGC50559
DTX3	BACE1
ACSF2	LAMB3
C11orf63	CACHD1
FBLN5	AK1
FABP4	MUC20
KIAA0831	KPNA5
HLF	transcript 3092561, GenBank L17325
MLLT3	TDH
DENND2C	IFT80
DAB1	SH3YL1
KLHDC1	TMEM98

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
KPNA5	AEBP2
CD9	IQCA
IRF6	SYTL2
LRRC37B2	MYO1B
MFAP4	CD9
IFT80	TMPRSS3
ZNF235	KIAA0408
CYP2B7P1	IGBP1
CEP97	SFTPC
TMEM106C	MAPK10
TWSG1	LOC400566
PALLD	FABP4
SSBP3	CLDN1
ABCA5	EPHX1
RAD50	PRKCQ
PER3	C6orf60
PRSS12	ERGIC3
CP110	B3GALT2
MARCH8	ULK2
C1orf198	CDH3
ZNF713	SDK1
ZER1	CNKSR1
CHRNB2	BTC
TMEM98	ALDH1A1
ZNF383	ST6GALNAC6
LOC51149	SLC4A4
EPHX1	LOC51149
C7orf31	BTBD9
RPL24	C2orf40
ERCC4	IKZF2
NIPSNAP3B	CELSR1
ZNF470	USP11
TRIB2	EPB49
C7orf58	CHCHD6
PLXNB1	HIBADH
PIGQ	SALL2
COQ10A	SUSD2
MYH10	HISPPD2A
NPAT	C1QTNF3

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
SAMD12	ATP2C2
EPB41L4A	PTPRF
CAMK2N1	AIG1
SYNE1	SLC4A8
FLJ22374	CRIP1
LMBRD1	FAM125B
TCTN2	BBS1
transcript 3184925, GenBank AK054857	PH-4
HMG20A	PTPN13
DGKG	EXPH5
MZF1	PLEKHB1
ZNF606	DLG3
H2AFV	SLC35A1
C8orf37	TPPP3
DSTN	TTC8
TGFB2	ZNF334
ZDHHC15	SNRPN
C9orf126	ZNF545
KALRN	FGFR2
ST3GAL4	TMEM116
C13orf15	CKB
STAM2	NELL2
TGFB3	C1orf101
EMP2	IRF6
YPEL1	CTSO
NPNT	SGSM2
HIBADH	MLLT3
LBH	JUP
TINAGL1	PLAC8
GUCY1A2	INADL
ZNF426	PSD3
C21orf63	NPAL3
CRIM1	CGNL1
RTKN2	FOLR1
PLA1A	LASS4
STXBP4	LRIG1
NMT2	KLC4
IFI27	FXYD3
PCDHB9	RGNEF

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
C16orf80	BTD
SCN7A	RCAN3
SOX5	EPHX2
GJC1	C8orf34
PLEKHH2	
LRIG1	
C9orf61	
ZNF680	
SLC4A8	
PGAP1	
BTC	
transcript 3092561, GenBank L17325	
N4BP2L2	
CCDC123	
ZNF180	
ZNF343	
TPM1	
CA3	
ZNF197	
C8orf42	
CCDC76	
GCNT4	
AIG1	
PXMP2	
LZTFL1	
ZBTB10	
ZFP2	
POGZ	
ZNF14	
CDRT4	
LRBA	
CTPS2	
C7orf36	
PLEKHF2	
NTN4	
ZNF567	
DPY19L4	
FAM122C	
ARGLU1	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
EMCN	
EFHC1	
KIF13A	
SH3RF1	
HBB	
RBL2	
INPP4B	
RAMP2	
GAS6	
IDS	
CKB	
TMEM116	
CEACAM6	
SALL2	
RAVER2	
FLJ11710	
AHSA2	
CACNA1D	
ODF2L	
MGAT5	
PPP2R5A	
HIST1H4B	
KRT8	
WHSC1L1	
DKK3	
RASSF9	
COX4I2	
PPAP2A	
ZFP3	
ZNF345	
NELL2	
BTBD16	
TTC21B	
CACNA2D3	
SPEF2	
SLIT2	
CCDC146	
MORN3	
CDC14A	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
PTPRM	
ZMAT1	
STX17	
ZNF571	
THEM4	
KALRN	
SHPK	
AMIGO2	
NDRG2	
NEGR1	
ALKBH8	
TMPRSS3	
MLLT4	
FOXJ1	
TNXB	
ATF7IP2	
PTN	
ZNF706	
transcript 2793198, GenBank AB062480	
PLK2	
ZNF75A	
MTERFD3	
ZNF721	
ABCC9	
C7orf41	
EPN2	
GPR56	
NR3C2	
GRIA1	
RPSA	
USP11	
TSPAN8	
EML1	
LAMA5	
GPC3	
C2orf13	
FLJ11996	
ADHFE1	
ZNF597	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
HSPA12B	
PAPD5	
ARHGAP6	
TXNDC13	
KAL1	
MIER3	
SVIL	
FLJ20160	
ABI2	
ZNF140	
BBS2	
PRDM5	
LOC401152	
KIF27	
ZEB1	
KLF12	
ZNF34	
ABCA6	
ZHX3	
CSPG4	
DMD	
HSPA2	
PDCD7	
DZIP3	
BPTF	
NEO1	
ESCO1	
TMEM47	
KLHL23	
PTPRS	
SERPING1	
PHLPP	
GPM6B	
C9orf68	
RGN	
TJP1	
PTPRD	
SCUBE1	
INADL	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
ERV3	
SHANK2	
KIAA1217	
PZP	
PPP1R14A	
RBJ	
CYP2B6	
WDR27	
SMARCC2	
CNTN4	
ADARB1	
PDZD2	
transcript 2363679, GenBank AK096078	
ZNF228	
C1QTNF7	
ZNF510	
FAM125B	
BBS5	
FABP3	
PRKAR2B	
LOC51336	
CD109	
MANEA	
TEK	
TSPAN6	
MUSK	
ABCA8	
ZNF83	
ZNF799	
C1orf102	
ZNF585B	
FRY	
GSTM2	
ZNF280D	
ZNF627	
KIAA0515	
STMN1	
LRRCC1	
FGFR3	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
TGFBR3	
LOC100130950	
SPATA6	
GABRB2	
KIAA1324	
CWF19L2	
ZNF792	
ALS2CR11	
GYPE	
CBX7	
CIRBP	
PARVA	
ZNF396	
TTBK2	
SUSD2	
FXYD3	
ZNF33B	
CLDN4	
HIF3A	
PHF3	
FARP1	
ZNF718	
ZNF329	
TNRC6C	
SPIN1	
CAPRIN2	
LARGE	
COL21A1	
APPBP2	
BBS4	
GSTA3	
TC2N	
OFD1	
LCA5	
ENPP5	
C1orf149	
PTGDS	
USP28	
KIAA0460	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
PAK3	
GRK5	
TDH	
WDR52	
GHR	
PVRL3	
AK3	
CFD	
TSPYL1	
SLC12A2	
NR1D2	
TMEM190	
KLF5	
TRIM56	
CDH6	
FANCL	
ZNF568	
SH3BP4	
KIAA1244	
SDK1	
LANCL1	
C6	
DIXDC1	
ADRB1	
ERCC6	
KIF2A	
PTPRF	
AGER	
NFATC3	
ANKH	
AFF2	
MAP2	
FERMT2	
C9orf52	
RNF182	
HSPC105	
SLCO2A1	
USP12	
NOX4	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
CRY1	
IQCK	
SLC25A23	
ARHGEF12	
ANGPT1	
WDR60	
WDR31	
GAB1	
APBB2	
ATP6V1E2	
GPR116	
DNAJB1	
ADH1B	
ZBTB4	
ZNF664	
MACF1	
PARD3B	
DNAH1	
BAIAP2L1	
PRKAA2	
SLC9A3R2	
DST	
FUZ	
MED6	
OLFML2A	
SLC16A12	
RPGR	
TSPAN12	
CAPS	
KCNRG	
FAM107B	
THSD4	
LTBP4	
CCND1	
ARHGAP29	
IL16	
CPA3	
ITGA8	
RGNEF	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
ZMYND11	
PARD6B	
UBTF	
ARHGAP5	
TMPO	
GATM	
ISLR	
AMOTL2	
NEK4	
LYRM5	
PSCD3	
CCDC117	
PER2	
FRS2	
SPTBN1	
GLCCI1	
VPS13A	
MYO5B	
NAB2	
KITLG	
CAP2	
C5orf4	
CTNND1	
DLEC1	
ZNF618	
SLC6A16	
DNAL1	
CABLES1	
BMPR1A	
RALA	
C1orf201	
ZNF544	
ZBTB39	
GEM	
ZBTB6	
KIAA0408	
FARS2	
ZNF208	
LRP11	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
YTHDC1	
SLC44A3	
SOD1	
PH-4	
THAP6	
TMEM107	
AQP3	
KIAA0423	
ZFP1	
PARD3	
VKORC1L1	
EPB41L2	
ZNF134	
TMEM67	
LOC201229	
CHIA	
PTPRK	
PLEKHA1	
SLC44A4	
ANUBL1	
RAB40B	
LRRC51	
WDR78	
USP54	
TMEM59	
SPOCK2	
SYNE2	
PUM2	
STXBP1	
PHF17	
ZNF337	
AKR1C2	
C2orf67	
PLCE1	
ADAMTSL3	
C10orf118	
GNA14	
KIF9	
ZBTB44	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
AKR1C3	
PRPSAP1	
DPT	
UBL3	
VEZF1	
SMURF2	
TFF1	
TAS2R60	
WISP3	
GABPA	
LAIR1	
GIP	
CTSB	
GLI1	
EBI3	
HCLS1	
CAMKK1	
SOX30	
WNT3	
C16orf24	
CTD-2090I13.4	
NHLH1	
CSF2RB	
SMAP2	
CHST11	
transcript 2772160, GenBank AF241539	
SPI1	
EPHA10	
TNFRSF1B	
NAGS	
SLC11A1	
TMSL1	
RCVRN	
RPS6	
CLEC11A	
CAMLG	
MS4A6A	
ACPT	
ADAMTS2	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
CALR	
UNQ6193	
OR2AK2	
FBXW9	
C5orf29	
SEMA6B	
TTYH3	
UNQ9419	
APOL5	
ATF5	
TREML2	
IGSF11	
CHRNB4	
C2orf7	
ZNHIT1	
KLK15	
KY	
C5orf27	
LILRB3	
MAGEC3	
SPATC1	
RETNLB	
SLC16A3	
MYH7	
CCL19	
ADAMDEC1	
CLRN3	
IFNW1	
DUPD1	
AQP7	
TSEN34	
KCNJ13	
HRH2	
C15orf27	
COL5A1	
GSX2	
ALDOAP2	
C21orf70	
TRIM7	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
EFHD2	
AMPD3	
SOLH	
SLC43A2	
GLDN	
GMIP	
CACNA2D4	
FAM71C	
PIK3AP1	
DARC	
HK3	
COX6A2	
SNX18	
ADORA3	
LY6G6C	
LILRB4	
ZNF541	
GLTSCR1	
HS3ST4	
RHOG	
ADAMTSL4	
PKM2	
OSR1	
C5AR1	
AURKC	
GMPPA	
CD163	
LOC642864	
C6orf81	
GRM3	
SIGLEC1	
TBC1D16	
MUSP1	
SPDYA	
AKR1D1	
C10orf90	
TMEM180	
GSTTP1	
TNC	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
ROR2	
PLTP	
FCGR2B	
LRRC15	
SIGLEC9	
VSIG4	
C12orf28	
LOC554234	
C20orf59	
NLN	
ISOC2	
PIK3R5	
NOD2	
TIMP1	
GAS7	
TG	
XDH	
LILRP2	
CACNA1E	
SLC38A7	
GALNT2	
TNFSF9	
SLC35E4	
RIN1	
IL1R2	
KLRG2	
EGLN1	
TRAPPC5	
OMG	
C16orf82	
MT1IP	
DSC2	
PFKFB3	
FCER1G	
SLC25A22	
MXD1	
FLVCR2	
SNX32	
TMEM171	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
TKTL2	
MAFB	
STAB1	
TXNDC17	
GLUL	
DSE	
NR5A1	
SIGLEC5	
LYN	
TCIRG1	
GRINA	
FLJ44006	
RNF149	
ACTN3	
GAFA1	
RHBDF2	
AGTRAP	
KCNQ5	
CES7	
GPR77	
CASQ1	
SHKBP1	
TPST1	
FLJ27255	
MPP1	
NPM3	
CTSL1	
OR1D2	
MMP3	
C3AR1	
CST9	
EPB41L3	
GPR84	
APRT	
EMR2	
TLR8	
ARRB2	
RNF166	
BOP1	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
ELOVL4	
C1orf162	
BAT1	
C4orf18	
EBI2	
TLR2	
C12orf34	
LRRC33	
ASGR2	
KIAA1949	
ORM2	
CLPP	
MGC13053	
MLLT1	
LILRB2	
SIRPB2	
NUP50	
AMPH	
RCN3	
CD300E	
AP3B2	
CLEC10A	
MTP18	
MRPL27	
NDP	
ORM1	
MERTK	
WAS	
SLC38A5	
LDLRAD3	
RGL1	
RAC3	
SIGLEC10	
PI3	
IGF1	
KCTD5	
DPYD	
C1orf38	
SLC22A16	

Human Genes Correlated with Firmicute Number	Human Genes Correlated with Proteobacteria Number
SAMSN1	
METTL7B	
OLFML2B	
MS4A6E	

Data Appendix 2 List of significant genes correlated with Shannon Diversity or OTU richness. Table relates to Chapter 4 on gene expression correlation to the bacterial microbiome.

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
FCGR2B	FCGR2B	EYA1	EYA1
EYA1	EYA1	TJP3	TJP3
PAX9	PAX9	CERKL	CERKL
	CDC20B	SEC14L3	SEC14L3
	RP11-738I14.8	CDC20B	CDC20B
	CCDC78	EPB41L4B	EPB41L4B
	MLL	C1orf168	C1orf168
	KRT5	CLGN	CLGN
	PTTG2	RIBC2	RIBC2
	CCDC108	ZMYND12	ZMYND12
	CCNE1	ALDH3A1	ALDH3A1
	SULF2	PAX9	PAX9
	SEC14L3	BTG4	BTG4
	EPB41L4B	FAT2	FAT2
	MAP3K14	KRT5	KRT5
	PSCA	DNAH2	DNAH2
	CERKL	CCDC78	CCDC78
	NUP50	KRT15	KRT15
	KCNRG	AK7	AK7
	CDK3	IGFBP2	IGFBP2
	FCGR3A	C6orf117	C6orf117
	FLJ45803	IFLTD1	IFLTD1
	C22orf15	TGM3	TGM3
	PTGS2	HSPA4L	HSPA4L

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	PI15	SLC22A23	SLC22A23
	ZNF660	RP11-738I14.8	RP11-738I14.8
	VWA3A	CLDN8	CLDN8
	MARCH1	STK33	STK33
	CYP2J2	FOXJ1	FOXJ1
	FAT2	CCNE1	CCNE1
	SLC22A23	FUZ	FUZ
	C21orf96	EHF	EHF
	DNAH2	ANKFN1	ANKFN1
	CXorf21	GSTA1	GSTA1
	SP110	VWA3A	VWA3A
	C1orf168	FLJ22167	FLJ22167
	CSF1R	C11orf70	C11orf70
	HRH1	IQCH	IQCH
	HIST1H3D	CCDC17	CCDC17
	C21orf59	DPY19L2P2	DPY19L2P2
	KRT15	LOC286187	LOC286187
	DBF4	CHST9	CHST9
	OR1L8	RABL5	RABL5
	FCGR2A	CXorf22	CXorf22
	CASZ1	FLJ40298	FLJ40298
	CCDC103	ALOX15	ALOX15
	TUBB2C	CASC2	CASC2
	ST6GALNAC2	DNAI2	DNAI2
	LRRC23	BAIAP3	BAIAP3
	AK7	ARMC4	ARMC4
	STK33	SLITRK6	SLITRK6

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	C2orf55	C10orf93	C10orf93
	CCDC67	ST6GALNAC2	ST6GALNAC2
	DYDC2	GOLSYN	GOLSYN
	KLK11	ARMC3	ARMC3
	KIF21A	WFDC1	WFDC1
	SLC8A1	CAPS	CAPS
	CCDC17	CCDC65	CCDC65
	DPY19L2P2	C3orf25	C3orf25
	CLGN	MMP21	MMP21
	UGT1A9	LRRC23	LRRC23
	WFDC1	DYNLRB2	DYNLRB2
	CHST9	PSCA	PSCA
	EHF	C11orf16	C11orf16
	ZMYND12	UGT1A9	UGT1A9
	ANKFN1	ALDH3B1	ALDH3B1
	MORN1	CASC1	CASC1
	RIBC2	C1orf87	C1orf87
	GSTA1	KIF6	KIF6
	C20orf165	CAPSL	CAPSL
	SLPI	LOC165186	LOC165186
	CLEC10A	CCDC135	CCDC135
	ZNF434	GLB1L2	GLB1L2
	GYS2	TUBB2C	TUBB2C
	STOX1	KLK11	KLK11
	NQO1	KIF21A	KIF21A
	ZNF436	TTLL6	TTLL6
	CCDC135	SLPI	SLPI

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	ATP12A	LOC200383	LOC200383
	IFITM3	CELSR1	CELSR1
	IFLTD1	PTPRT	PTPRT
	ARMC4	RAB36	RAB36
	ALDH3A1	APOBEC4	APOBEC4
	LIN37	YSK4	YSK4
	KNDC1	DNAI1	DNAI1
	PQLC3	ZMYND10	ZMYND10
	KIF6	CCNA1	CCNA1
	C6orf150	MORN2	MORN2
	FOXJ1	CETN2	CETN2
	CCDC40	DNAH10	DNAH10
	CAPS	FRMPD2	FRMPD2
	CXorf22	FAM81B	FAM81B
	FLJ40298	PACRG	PACRG
	SOX2OT	CCDC67	CCDC67
	SLC6A4	DYDC2	DYDC2
	GOLSYN	C1orf173	C1orf173
	TJP3	RSPH1	RSPH1
	C6orf117	C14orf50	C14orf50
	LOC100129540	RP11-529I10.4	RP11-529I10.4
	CELSR1	ATP12A	ATP12A
	LOC376693	C2orf39	C2orf39
	DNAJA4	NEK5	NEK5
	C8orf4	IQUB	IQUB
	DNAI2	PHGDH	PHGDH
	C1orf87	DCDC5	DCDC5

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	HSPA4L	FLJ46266	FLJ46266
	PLLP	RTDR1	RTDR1
	IL1R2	DNAH3	DNAH3
	C1QA	APOBEC3G	APOBEC3G
	BAIAP3	FCGR2A	FCGR2A
	KCND2	RUVBL2	RUVBL2
	ADCY7	C20orf26	C20orf26
	RCAN1	DNAH10	DNAH10
	PRPF40B	ANKRD54	ANKRD54
	CASC2	C6orf118	C6orf118
	VIPR1	FGF14	FGF14
	C9orf9	RGS22	RGS22
	CCDC65	PDE4B	PDE4B
	OBFC2A	CXorf21	CXorf21
	RGS2	RHOH	RHOH
	XAF1	KIAA0319	KIAA0319
	C10orf93	TMC4	TMC4
	GPX2	SPA17	SPA17
	FLJ22167	TRIM29	TRIM29
	LAMA1	CCDC60	CCDC60
	DYX1C1	LRRC34	LRRC34
	HSPA6	TSPAN6	TSPAN6
	CLDN8	AKAP14	AKAP14
	SIGLEC1	C1orf102	C1orf102
	NCF4	transcript 3683871, GenBank AK027211	transcript 3683871, GenBank AK027211
	ARID5A	KRT19	KRT19
	C11orf70	SPATA17	SPATA17

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	CCNA1	LRGUK	LRGUK
	SLC20A2	C11orf66	C11orf66
	CETN2	OBFC2A	OBFC2A
	SLC9A9	C10orf63	C10orf63
	IKZF1	CCDC33	CCDC33
	ARMC3	NEK10	NEK10
	APOL4	C2orf55	C2orf55
	PACRG	KCNRG	KCNRG
	STAB1	ANKRD22	ANKRD22
	CAPSL	C19orf51	C19orf51
	CLCA2	APOL4	APOL4
	SOCS1	DCDC1	DCDC1
	FUZ	C1orf88	C1orf88
	TSPAN6	CCDC103	CCDC103
	C14orf50	GSTA2	GSTA2
	FLT3	DNAH7	DNAH7
	CCDC81	EFCAB1	EFCAB1
	CA11	CCDC81	CCDC81
	C1orf110	LRRIQ1	LRRIQ1
	MUC15	SLC44A4	SLC44A4
	HRH2	FCGR3A	FCGR3A
	LILRB1	ASB14	ASB14
	SLC2A9	C22orf15	C22orf15
	TACC2	EFCAB6	EFCAB6
	G6PC2	SP110	SP110
	SLC27A2	LIN37	LIN37
	HSPBP1	KNDC1	KNDC1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	TMEM176A	PPP5C	PPP5C
	BNIP3L	TNFSF13B	TNFSF13B
	ESRRA	MLL	MLL
	C20orf26	C1orf110	C1orf110
	DCDC1	ZNF474	ZNF474
	GPR87	WDR69	WDR69
	LUM	CYP2J2	CYP2J2
	C2orf39	ROPN1L	ROPN1L
	PTPRT	C21orf59	C21orf59
	RAB36	FLJ23834	FLJ23834
	STC1	RORC	RORC
	PDE4B	PLLP	PLLP
	HAO1	PIH1D2	PIH1D2
	RUVBL2	DNAJA4	DNAJA4
	CLDN16	CDS1	CDS1
	FRMPD2	C10orf107	C10orf107
	CCDC60	LRRC43	LRRC43
	NDRG4	LRRC48	LRRC48
	SMAD5OS	SLC9A11	SLC9A11
	DCDC5	DNAH5	DNAH5
	RP11-529I10.4	C1orf158	C1orf158
	CMTM4	C6orf165	C6orf165
	PRRX1	C21orf96	C21orf96
	FLJ45121	RICS	RICS
	KRT79	GPR87	GPR87
	C1orf173	SLC27A2	SLC27A2
	KIAA1199	CXorf41	CXorf41

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	ARNT2	DYX1C1	DYX1C1
	MIPEP	FCGR2B	FCGR2B
	CASC1	DUOX1	DUOX1
	FOXA1	C10orf79	C10orf79
	FLJ46266	TEKT2	TEKT2
	DNAI1	TEKT3	TEKT3
	MAGEF1	GAS2L2	GAS2L2
	BHLHB5	KLHL32	KLHL32
	SDC3	RAGE	RAGE
	TEKT2	CYP4X1	CYP4X1
	SCUBE2	SOX2OT	SOX2OT
	LEKR1	KIAA1324	KIAA1324
	SEC14L2	PFN2	PFN2
	C1orf88	TACC2	TACC2
	RASSF4	VWA3B	VWA3B
	LRRIQ1	CASZ1	CASZ1
	LOC286187	NTF3	NTF3
	GLB1L2	DSP	DSP
	ANKRD22	KATNAL2	KATNAL2
	DNAH3	EFHC2	EFHC2
	CTHRC1	DNAH12L	DNAH12L
	TIAM1	WDR65	WDR65
	DNAH7	C14orf179	C14orf179
	SPA17	HSPBP1	HSPBP1
	DNAH10	C20orf96	C20orf96
	UGT2A1	C9orf116	C9orf116
	SCO2	PRDX5	PRDX5

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
	RABL5	MORN1	MORN1
	RGS22	WDR16	WDR16
	PRND	ANKRD35	ANKRD35
	C3orf25	MUC15	MUC15
	AKAP14	FBXO15	FBXO15
	DOC2A	CCDC13	CCDC13
	CD74	ZNF660	ZNF660
	IQUB	IL1R2	IL1R2
	DUOXA1	C14orf45	C14orf45
	RORC	PTTG2	PTTG2
	ROPN1L	NUP62CL	NUP62CL
	ADORA2A	TCTN1	TCTN1
	C19orf51	DNAJB13	DNAJB13
		MS4A8B	MS4A8B
		TTN	TTN
		PTPRU	PTPRU
		C12orf63	C12orf63
		CCDC11	CCDC11
		ABHD11	ABHD11
		transcript 3973556, ENSEMBL Prediction GENSCAN00000028495	transcript 3973556, ENSEMBL Prediction GENSCAN00000028495
		MIPEP	MIPEP
		AGBL2	AGBL2
		FAM50B	FAM50B
		C20orf85	C20orf85
		MDH1B	MDH1B
		SPAG1	SPAG1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
		NUP50	NUP50
		ANKRD45	ANKRD45
		DUOXA1	DUOXA1
		FANK1	FANK1
		TPH1	TPH1
		ZBBX	ZBBX
		C9orf9	C9orf9
		TEX9	TEX9
		KLHDC9	KLHDC9
		SPAG16	SPAG16
		CLCA2	CLCA2
		FOLH1	FOLH1
		CDON	CDON
		C10orf81	C10orf81
		TTC18	TTC18
		C6orf206	C6orf206
		IFT172	IFT172
		KCNE1	KCNE1
		DYDC1	DYDC1
		SLC22A4	SLC22A4
		KIF9	KIF9
		C9orf24	C9orf24
		C9orf98	C9orf98
			OSBPL6
			RELT
			CAPS2
			HUNK

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			MARCH10
			STOML3
			NME5
			SERPINI2
			DNAH9
			FAM13A1
			KRT79
			NQO1
			SPAG6
			SOCS1
			SLC2A9
			FAM92B
			C9orf68
			HOMER2
			NSUN7
			CCDC113
			LRRC46
			MKS1
			TTC29
			WDR49
			SULF2
			C17orf87
			LOC100128751
			SPAG17
			LRRC50
			C3orf67
			GLYATL2

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			ARMC2
			CCDC148
			CCDC40
			NDRG4
			CMTM4
			PLEK
			PROM1
			LCA5L
			ARID5A
			WDR31
			WDR78
			MORN3
			TIAM1
			FAM3D
			NEK11
			DNAH11
			PLEKHB1
			ZEB2
			SCGB3A1
			WDR66
			HHLA2
			LEKR1
			CCDC108
			RP1
			RABL2B
			OR1L8
			ESRRG

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			RFX3
			UBXD3
			SPATA18
			CCNO
			LRP11
			DLEC1
			THSD4
			EFHB
			TSPAN5
			KCTD1
			CDK3
			RFX2
			ARHGAP30
			BMPR1B
			C6orf97
			SLC8A1
			DNHL1
			ADORA2A
			SLC20A2
			LOC389118
			CCDC37
			HYDIN
			ALS2CR12
			FOXA1
			WDR63
			KLF5
			ZDHHC1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			MARCH1
			VTCN1
			HELB
			TTC16
			TPM2
			EZR
			BCAS1
			CHST6
			CA11
			C15orf26
			ADCY7
			CLEC9A
			SLC25A4
			C9orf18
			CD86
			CLDN7
			SRGAP3
			GSTA3
			INDOL1
			HEXIM2
			ST6GALNAC1
			FLT3
			TRAF3IP1
			NCKAP1L
			C8orf4
			TEKT1
			HSPB2

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			RCAN1
			C13orf30
			TCTN2
			FLJ37464
			ADH7
			AQP3
			FLJ45803
			NEDD4L
			LRRC51
			PCSK2
			FAM81A
			C6orf150
			ZNF440
			MMP9
			ALCAM
			DNAL1
			FLNB
			VSTM2L
			TPPP3
			GLRX
			PDK1
			LOC376693
			C10orf64
			LAX1
			GPX2
			SPATA4
			C3orf15

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			TMED6
			NEK10
			HRH1
			DOK2
			DOCK2
			ICA1L
			ITGB4
			CLEC10A
			KRT8
			TAGLN
			GPR162
			ZNF436
			TRIM38
			TMEM156
			NCF4
			RABL4
			C10orf92
			SCNN1A
			TSPAN1
			WDR93
			NPR1
			APOBEC3D
			DMKN
			MAT1A
			SERPINB4
			GRHL1
			CYBA

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			TTC26
			TUBA4B
			ZNF434
			INPP5D
			RIBC1
			LOC100129540
			TAS2R40
			PSAT1
			NME7
			COL4A5
			UNQ5814
			ASNS
			TMEM45B
			KIF24
			NPHP1
			GRHL2
			RBM24
			CLDN16
			TP63
			DOC2A
			ABCA1
			ACTN4
			B9D1
			CCRK
			CCDC108
			FAM72A
			UGT2A1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			CLEC7A
			SH3BP1
			HGF
			DOCK8
			CDC42EP3
			MOXD1
			BTN2A3
			DENND1C
			GPR20
			SLC2A3
			REEP1
			TNFAIP2
			FSIP1
			C9orf6
			TSNAXIP1
			CD69
			SRD5A2
			IQCK
			PIK3CG
			LRWD1
			PTGS2
			FSD1L
			CD40
			CD19
			SPATS1
			ATP1B1
			FKBP11

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			CTTN
			TTC25
			IGLV6-57
			ADAM19
			PTPRC
			DMGDH
			KIF19
			CLDN4
			TPPP
			TMEM190
			GSTO2
			SULT1B1
			IKZF1
			PTAFR
			POU2F2
			LUM
			SLC9A9
			KALRN
			STK36
			GYS2
			TLCD1
			FBLN2
			UBAC1
			SLC6A4
			MAP1A
			TRIP13
			C1orf201

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			CYP1B1
			KRT23
			ACYP1
			CLMN
			DEGS2
			PLEKHG7
			BLK
			LRRC27
			EFHC1
			CCDC114
			MAP3K14
			C4orf22
			CAP2
			ABBA-1
			VIPR1
			MDFIC
			DTX3
			PRUNE2
			GBP5
			DSC3
			CCDC57
			TJP2
			SLC23A1
			MGC2752
			UPP1
			FBXO16
			SLAMF1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			SLC44A5
			RCSD1
			IFT140
			FLJ25439
			RUVBL1
			KIAA1199
			LILRB1
			VCAM1
			SPEF2
			ANKRD18A
			PLXNB1
			IFI44L
			IFT81
			FLJ21511
			C1orf92
			CCDC19
			TMEM17
			HIST1H3D
			GAL3ST4
			YAP1
			CSF1R
			DIRAS1
			MS4A14
			CNN1
			PEG10
			XAF1
			FCRLA

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			CTLA4
			SP140
			TTC12
			TACSTD2
			IL2RG
			LPAR3
			IGJ
			C1orf34
			TRIM14
			KIAA0125
			DVL1
			IL5RA
			APOO
			PSCD4
			HAL
			SCO2
			CDC7
			N6AMT1
			LOC400566
			CDKL3
			KCTD5
			MYEF2
			ABCC8
			IL1B
			LDHC
			LRP2BP
			SIRPD

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			TSGA10
			ACTR3B
			IRF8
			CHRM1
			VAV1
			MAPRE3
			TMEM67
			GPR141
			SERPINB11
			BHLHB5
			CPVL
			CUEDC1
			FAP
			IFFO
			DEFB1
			SAE1
			TM4SF19
			CCR2
			PRND
			MMP2
			VSIG1
			MAPK15
			ARNT2
			PIK3C2G
			USP2
			POPDC3
			transcript 3739859, GenBank AF269286

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			PQLC3
			CD164L2
			PI15
			TRIM69
			TRIM2
			SCUBE2
			LPAL2
			PIGU
			LCN2
			MCAT
			RNASET2
			BTLA
			MAP4K1
			SPEF1
			P2RY13
			WDR13
			DPY19L2P2
			PIK3CD
			FXC1
			PTPN3
			BEST4
			RNF43
			CSF2RB
			CDKL5
			HS3ST3B1
			ADAM12
			KCND2

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			OR4D10
			PLD2
			SAMD3
			ASPHD2
			PEBP4
			LOC400451
			C10orf57
			PKP1
			CLDN3
			C20orf12
			PLB1
			MYL9
			TMEM206
			DUSP6
			ACSBG2
			LAT
			NCF2
			CYP2B6
			MEI1
			RGS2
			SHROOM3
			GSTP1
			LRRC39
			GUCY1B2
			TBX3
			IFITM3
			EPSTI1

Shannon Diversity (FDR <0.1)	Shannon Diversity (FDR <0.25)	Total Bacterial Species (FDR <0.1)	Total Bacterial Species (FDR <0.25)
			MAPK10
			KCNJ16
			MUC20
			NFX1
			ARHGAP15
			OSM
			ENOSF1
			ANKMY1
			C10orf67
			FCAR
			CCDC146
			NBEA
			WDR54
			BACH1
			PIK3AP1
			LLGL2
			SRGN
			ALOX5