Multi 'Omics Integration of the HIV Airway Epithelium: Integration of the Microbiome, Transcriptome and Methylome

by

Marcia Smiti Jude

B.Tech., Anna University, 2019

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

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Multi 'Omics Integration of the HIV Airway Epithelium: Integration of the Microbiome, Transcriptome and Methylome

submitted by Marcia Smiti Jude in partial fulfillment of the requirements for

the degree of Master of Science

in Experimental Medicine

Examining Committee:

Dr. Janice Leung, Clinical Assistant Professor, Medicine, UBC
Supervisor

Dr. Don Sin, Professor, Medicine, UBC
Supervisory Committee Member

Dr. Christopher Carlsten, Professor, Medicine, UBC
Supervisory Committee Member

Abstract

RATIONALE: People living with HIV (PLWH) appear to have increased proneness to chronic obstructive pulmonary disease (COPD) independent of their cigarette smoke exposure. Previous studies have shown that HIV infection is associated with changes in the airway microbiome and host response, however the exact mechanism of disease progression is still unknown. We hypothesize that airway epithelial dysbiosis in PLWH increases the susceptibility to COPD in this group.

METHODS: Airway epithelial cell brushings were obtained from 18 COPD+HIV+, 16 COPD-HIV+, 22 COPD+HIV- and 20 COPD-HIV- subjects. Microbiome, methylation and transcriptome profiles were measured using 16s amplicon sequencing (Illumina Miseq®), Illumina Infinium Methylation EPIC chip®, and RNA sequencing (NovaSeq 6000®), respectively. Microbiome analysis was performed using QIIME 2™, and transcriptome and methylation analyses were performed using R language. The three datasets were integrated using Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) implemented in the mixOmics R package. Fifty repeats of 10-fold cross-validations and a correlation threshold of 0.7 were set to determine key interactions between bacterial ASVs, CpG methylation sites, and gene transcripts amongst the subjects based on their COPD, HIV and combined COPD and HIV statuses.

RESULTS: The microbiome analysis identified that the groups most associated with disease (COPD+, HIV+ and COPD+HIV+ groups) had reduced alpha diversity (Shannon Diversity Index p=0.0013, p=0023 and p=0.0002, respectively), and significantly disrupted microbial communities (Bray Curtis PERMANOVA p=0.001, p=0.007 and p=0.001, respectively) compared to their relatively "healthy" counterparts. This was accompanied by changes in the host transcriptome and epigenome, our analysis of which identified top genes and CpG sites that were differentially regulated in patients with COPD and/or HIV. Integration of the three -omes identified features that were correlated with one another at a threshold >0.70. On combining the COPD and HIV statuses of subjects, the multiomic integration identified correlations between the bacterial ASV *Bacteroidetes Prevotella* and transcriptomic features *FUZ*, *FASTKD3* and *ACVR1B*, and epigenetic features *CpG-FUZ* and *CpG-PHLDB3*. It may be that these features together influence host pathways regulating mucociliary clearance, respiration and energy, cell cycle, and immunity.

Lay Summary

People living with HIV are increasingly susceptible to COPD independent of well-known risk factors like tobacco smoking. This increased respiratory burden has been attributed partly to lung microbiome alterations, which may be linked to changes in host response. Our microbiome analyses validated previous reports that there is reduced microbial diversity and community structures changes in COPD and/or HIV patients compared to healthy individuals. This may be accompanied by changes in host processes such as gene expression and methylation, our analysis of which identified top were differentially regulated genes and methylation sites, respectively, in patients with COPD and/or HIV. On integrating the microbiome, transcriptome and methylome, we identified highly correlated features that may be targets against HIV-associated COPD. We especially highlight the microbial feature *Bacteroidetes Prevotella* which was highly correlated with genes and methylation sites related to host pathways such as cilium assembly, mitochondrial respiration, cell cycle regulation and inflammation.

Preface

This research was approved by the Providence Health Care Research Institute ethics committee (Certificate H15-02166).

I performed all the data analysis and wrote the entire first draft of the manuscript. Subsequent drafts were edited by Dr. Janice Leung. Dr. Fernando Studart assisted with the microbiome analysis. Statistician Chen Xi Yang assisted with the transcription and methylation data analysis.

The research forming this thesis has been published as abstract as part of the ATS 2020 International Conference.
Jude, M. S., Yang, C. X., Leitao Filho, F. S. S., Yang, J. S. W., Shaipanich, T., Kobor, M., ... & Leung, J. M. (2020). Multi 'Omics Profiling of the HIV Airway Epithelium: Integration of the Microbiome, Methylome, and Transcriptome. In B28. HOST AND MICROBIAL CLINICAL STUDIES IN LUNG INFECTIONS AND LUNG DISEASES (pp. A2940-A2940). American Thoracic Society.

Table of Contents

Appendix

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest appreciation and gratitude to my research Supervisor Dr. Janice Leung for their unparalleled support, patience, meticulous scrutiny, and constructive and scholarly advice. Thank you for challenging my knowledge, guiding experiments as my thesis developed, and helping me keep perspective on where my research fits into the bigger picture and its relevance clinically.

I would also like to extend my gratitude to friends, colleagues and mentors at the Leung lab and at the Centre for Heart Lung Innovation, St. Paul's Hospital and UBC for their relentless support and helpful advice.

I would also like to thank members of my Supervisory Committee Drs. Don Sin and Christopher Carlsten for their keen interest in my research and invaluable insights.

Special thanks are owed to my family, who have supported me throughout my years of education, both morally and financially, and my friends who have motivated me throughout my graduate school experience.

Chapter 1

Introduction

1.1 Chronic Obstructive Pulmonary Disease (COPD)

1.1.1 Background and Significance

The World Health Organization projects COPD to become the third leading cause of death by 2030 [151]. COPD is a major public health problem, with morbidity and mortality rates on the rise. The lack of consensus on aspects like screening, staging, assessment and treatment of the disease deem it a formidable challenge for healthcare systems worldwide[193].

COPD is defined by the Global Initiative for Obstructive Lung Disease (GOLD) as "a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases"[69]. It is denoted by two frequently coexisting conditions: small airways disease and emphysema. Small airways disease is marked by inflammation and destruction of the terminal bronchioles, resulting in mucus hypersecretion by goblet cells and airway obstruction[96]. In emphysema, alveolar walls are damaged leading to loss of elasticity of air sacs, making it difficult for people with the disease to fully expel air out of their lungs. Additionally, emphysema can cause the alveoli to rupture reducing the surface area available for gas exchange[70].

1.1.2 Risk Factors

There are a number of factors that play a role in COPD, with tobacco smoking being the most commonly encountered risk factor. Tobacco contains a number of noxious components that can cause abnormal pulmonary inflammation and the release of pro-inflammatory mediators by inviting the host's immune system to attack the lung tissue. This causes severe oxidant/antioxidant imbalance, protease/anti-protease imbalance, dysfunction of the autonomic nervous system, and changes in cholinergic nerves that further aggravate pulmonary inflammation, all resulting in extensive lung tissue destruction[7]. The pivotal role of smoking as a risk factor can also be explained by the fact that smoking cessation has been shown to delay the onset of COPD symptoms and improve quality of life in patients[106]. However, only a small percentage of chronic heavy smokers show decline in lung function and develop COPD, which highlights the contribution of other risk factors to the disease

[174]. These include environmental and occupational exposures to dusts, fumes and indoor biomass fuel burning, and repeated respiratory infections. There are also known genetic risk factors such as alpha 1-antitrypsin deficiency that can contribute to the development of COPD [175].

Recently, patients living with human immunodeficiency virus (HIV) have also been shown to have an elevated risk for COPD [49]. The contribution made by the virus towards the increased burden of obstructive lung disease is further complicated by higher rates of smoking, illicit drug use, previous pulmonary opportunistic infections and low socioeconomic status seen in the HIV-infected population [100] [128]. Despite that, a number of studies suggest that HIV is an independent risk factor for the development of COPD [137] [5].

1.2 Human Immunodeficiency Virus

Human Immunodeficiency Virus (HIV) is a virus that attacks the body's immune system. It is a member of the Lentivirus genus of the Retroviridae family. HIV isolates are grouped into two main types, HIV-1 and HIV-2, of which the most common and infectious strain is HIV-1. HIV infection is a chronic, potentially life-threatening condition, typically with a long period of clinical latency and persistent viral replication [58].

1.2.1 Immunopathogenesis of HIV Infection

HIV is transmitted via certain bodily fluids such as blood, semen, vaginal secretions, and breast milk, when they come in contact with mucous membranes, damaged tissue or via direct injection into the bloodstream [188]. In fact, HIV cannot survive outside the host's bloodstream or lymphatic tissue [150].

A multitude of processes are involved in the establishment and progression of HIV disease. Upon primary infection, HIV-infected cells can transfer the virus to immune cells such as T-cells, macrophages,and dendritic cells, as well as cells lining vaginal or anorectal mucosae [129]. HIV enters dendritic cells, migrates to the lymph nodes and disseminates to its primary target cells, CD4+ T-cells. Viral entry into T-cells is mediated by the interactions between the envelope glycoprotein gp120 and the CD4 receptor and a co-receptor (mainly the chemokine receptors CCR5 and CXCR4) [115]. Infected cells, especially macrophages and resting CD4+ T-cells, can act as viral reservoirs and establish latent infection by escaping the viral immune response and effective regimens of antiretroviral therapy. Given appropriate stimulus, they can reactivate and start producing infectious virions [2]. In addition to causing an aberrant activation of the immune system, HIV also increases levels

of certain proinflammatory cytokines which regulate viral expression in other tissues [208].

In the early stages of infection (often referred to as "acute infection" and occurring in the first two weeks), plasma viral levels rise exponentially to reach a peak of over 10^6 RNA copies/ml [17] [168]. The symptoms of acute HIV infection that typically first present during this phase include fever, fatigue, rash, headache, and enlarged lymph nodes. These symptoms resolve spontaneously after 1 to 2 weeks, following which patients can remain in a latent asymptomatic state for years [132]. Because the symptoms of HIV are initially nonspecific, acute HIV infection is often underdiagnosed or misdiagnosed as a variety of other illnesses, including infectious mononucleosis, secondary syphilis, acute infection with hepatitis A or B, roseola or other viral infections, and toxoplasmosis [93].

The acute infection stage is characterized by high levels of infectivity, and is normally short-lived because the host's immune response responds to and regulates viral replication. After the initial rise in plasma viremia, the titers decline over a period of months until they reach a steady-state level of viral replication (viral set-point) or drop below detection level [93] [120]. The viral set-point may be influenced by factors related to genetic differences in co-receptors [191] [83], qualitative differences in the immune response [162], or differences in the virulence of viral strains [99]. Establishing a lower set-point, as a consequence of lowering the viral load during primary infection, is clinically important as it slows disease progression.

With the drop of HIV viremia levels and absence of symptoms, most infected persons enter into a clinical asymptomatic period. During this period, the host's immune system recognizes viral antigens present on surfaces of infected cells and promotes their elimination by antigen-specific cytotoxic mechanisms [9], while the virus continues to replicate at the reservoir lymph sites. Several factors are implicated in why antiviral immunity is unable to completely eradicate the infection. These include the persistence of virus in the reservoir sites, low expression of viral antigens, and the high rate of mutations within the viral genome. The resulting dynamic equilibrium between HIV replication and host antiviral immunity sets the stage for chronic systemic inflammation. Continuous immune activation may arise due to the presence of HIV, microbial products and co-infections, and other homeostatic mechanisms [61]. This milieu allows for further viral replication, destruction of the lymphoid tissue architecture, repeated cycles of T-cell loss and replenishment, eventually culminating in the functional exhaustion of T-cells and immunodeficiency [76].

The further progression of the disease to AIDS is dependent on the capacity of the host to contain viral replication and to reconstitute the pool of memory T-cells. However, as the disease advances, CD4 T-cell numbers decline gradually (< 200 cells/µl) and the immune system is impaired, creating permissive conditions for various parasitic [19], bacterial, viral and fungal co-infections [60], tumours, and other complications. Infected individuals expe-

rience severe reduction of body weight, fever, respiratory and gastrointestinal symptoms, night sweats, and oral or genital ulcers. During the AIDS phase, encephalopathy [30], anaemia and marked lymphopenia [159], neoplastic diseases such as Kaposi's sarcoma and lymphomas [177] are frequently detected. The progression of the HIV disease is extremely variable, depending on viral control and the antiviral response of the host. Most patients diagnosed with HIV are likely to develop AIDS within ten years if left untreated [212]. However, if antiretroviral therapy is initiated, the individual may achieve lifespans comparable to the general population [127].

Antiretroviral therapy (ART), consisting of a combination of antiretroviral drugs, is now the standard-of-care therapy for people living with HIV (PLWH) [152]. Although it is not curative, ART can provide longer lives for patients and reduce HIV transmission by suppressing HIV replication in infected cells and lowering the plasma viral load [95] [153].

1.3 HIV and Lung Function

Studies conducted in the pre-ART era first noted the association between respiratory function abnormalities and HIV infection [167]. During this era, the pulmonary manifestations of HIV were dominated by infectious complications [85]. PLWH exhibited reduced diffusing capacity for carbon monoxide (DLCO), increased emphysema, increased airway obstruction and small airways disease, and increased prevalence of dyspnea, cough and sputum production than HIV-negative subjects. Risk factors for respiratory symptoms and pulmonary function abnormalities in the pre-ART era included smoking, illicit drug use, low CD4 counts and pulmonary infections. [167] [68] [41][134][43].

In the ART-era, pulmonary complications, especially COPD and asthma, are common in PLWH, particularly in those who smoke [205][79]. In cohort studies of PLWH, it was noted that 16–20% have asthma or COPD, 21% have obstructive ventilatory defects, and more than 50% have reduced diffusing capacity measurements [49]. Another study by Drummond *et al* on PLWH who concurrently inject drugs showed that with poorly controlled HIV infection (CD4 cell count < 100/uL), FEV1 decline is accelerated by 57 mL/year compared to HIV-uninfected patients [51]. Crothers *et al*, in a prospective observational study of HIV-infected and uninfected men enrolled in the Veterans Aging Cohort 5 Site Study identified that HIV infection was an independent risk factor for COPD [34]. In a later study, they analyzed data from age, sex, ethnicity and site-matched HIV-infected and uninfected veterans, and observed that the HIV-infected were more likely to develop COPD, asthma, as well as pulmonary infections, when compared to the uninfected group. This greater burden of pulmonary complications in HIV-infected patients was attributed to aging-related changes in lung health [35] .

1.3.1 Mechanisms of COPD in HIV

The unique mechanisms for the increased incidence of non-opportunistic chronic lung diseases in the milieu of a viral infection are poorly understood. However, there are a number of hypotheses proposed to understand the pathogenesis of COPD in HIV [105]. The prevalence of low socioeconomic conditions and risky behaviors like injection/inhalational drug use and tobacco smoking, particularly high in PLWH, are factors that may be responsible for the increased risk of COPD observed in this population [72] [148]

Some studies conducted in the ART-era have also reported ART to be an independent predictor of increased airway obstruction [64] [68]. The exact mechanism of how this happens is unknown, but potential explanations include: 1) the direct effects of ART - it is proposed that antiretroviral agents, particularly protease inhibitors, may cause endothelial damage to the pulmonary capillary bed resulting in the reduced effective blood volume for gas exchange (reduced DLCO) [42]; 2) immune reconstitution inflammatory syndrome - restoration of the immune system after ART is initiated might result in a state of chronic inflammation propagated by colonizing organisms, autoantigens, or by HIV itself [138][74]; 3) possible development of auto-immunity - which may occur as a side effect of restoring immunocompetence in previously immunocompromised individuals after the successful introduction of ART [169]. Nonetheless, a recent randomized substudy in the Strategic Timing of Antiretroviral Treatment (START) cohort identified no significant difference in lung function decline between those that started ART immediately vs. those that deferred until CD4+ T-cell counts were 350 cells/mm^3 or AIDS developed [104].

Inflammation associated with sub-clinical infections has also been proposed as a risk factor for the development of COPD. *Pneumocystis* pneumonia (PCP), a cause of pneumonia in immunocompromised hosts, is a leading cause of death in HIV-infected individuals. Although, with the development of ART, the incidence of PCP has reduced, it is still a cause for concern in individuals with no access or poor response to ART [139]. *Pneumocystis* has been implicated in the "vicious circle" hypothesis where colonization by the organism particularly in the lower airways perpetuates a state of inflammation and lung tissue remodelling [183]. Evidence has also linked *Pneumocystis jirovecii*, a fungal opportunistic pathogen, to COPD development and severity in HIV-infected and uninfected individuals [142] [141] [24] [81]. Shipley *et al* used cynomolgus macaques infected with chimeric simian-human immunodeficiency virus (SHIV) as an AIDS model to study *Pneumocystis* colonization in HIV-associated COPD. They recorded significant emphysematous tissue damage and elevated levels of pro-inflammatory mediators in the BAL of SHIV-infected macaques with *Pneumocystis* colonization, and minimal effects in macaques without *Pneumocystis* [186].

Increased oxidative stress is another potential mechanism linking HIV and COPD. COPD,

largely being a disease of the elderly, is associated with accelerated ageing and oxidative stress [10]. This may occur in part, as a result of exposure to cigarette smoking and other pollutants, but also from the activation of inflammatory cells, particularly during HIV infection, which produce reactive oxygen species (ROS) [138]. This could explain the characteristic signs of accelerated ageing such as telomere shortening, DNA damage, mitochondrial dysfunction, reduced autophagy, accelerated cellular senescence and death, seen in these patients [214].

Epidemiological data suggest that HIV itself may be an independent risk factor for COPD, after adjusting for age, ethnicity and pack-years of smoking [34]. Uncontrolled HIV infection, with lower nadir CD4 cell count and higher viral load, may be associated with an increased risk for obstructive lung diseases, including COPD and asthma [50] [173] [49]. HIV-infected individuals also have an increased susceptibility to bacterial colonization, aberrant inflammatory responses, altered oxidant–antioxidant balance, increased apoptosis, decreased respiratory muscle function, systemic effects of HIV-related viral proteins, and a host of other factors that may play a role in the accelerated lung function decline [194][105]. Many of these factors may act together and result in the many manifestations seen in the HIV lung [138].

Chapter 2

HIV and the Lung Microbiome

2.1 Introduction

2.1.1 What is the Microbiome?

The term 'microbiome' refers to a specific biological niche, including the assemblage of microorganisms (also called 'microbiota'), their genomic content and metabolic products, and the surrounding environmental conditions [126]. The microbiome is a living ecosystem, including, but not limited to, a collection of microorganisms such as bacteria, archaea, viruses and fungi, which form a highly complex network of interactions between each other and the host. The composition of the microbiome is unique in each individual, and is influenced by factors such as body site, diet, antibiotics, lifestyle, socioeconomic status, pollution and other environmental factors [67]. Most elements of these microorganisms inhabit the human skin, nails, eyes, genitalia, oral and upper respiratory and gastrointestinal tracts, and are harmless in healthy individuals [121].

Due to its influence on human health and disease, the microbiome has especially been of research interest recently. A review conducted by Petersen *et.al.* characterized changes in the microbiome (or dysbiosis) as the (i) loss of beneficial microbial organisms, (ii) expansion of potentially harmful microorganisms and/or (iii) loss of overall microbial diversity. Microbial dysbiosis has been shown to have overall effects on health, immunity, development, and disease progression [156]. In recent years, microbial dysbiosis has been implicated in numerous diseases, including inflammatory bowel disease [65] [62], multiple sclerosis, type-1 diabetes [221] [222], allergies [206], asthma [78], autism [59] [80], and cancer [102] [63].

2.1.2 The Respiratory Microbiome, COPD and HIV

The lung microbiome differs between individuals with and without COPD [84] [37]. Studies by Hilty *et al* with bronchial brushings found increased levels of *Proteobacteria* (particularly *Haemophilus* spp.) and reduced levels of *Bacteroidetes* (particularly *Prevotella* spp.) in asthma and COPD patients compared to controls [78]. Extending these findings in lung tissue samples, Sze and colleagues observed that the genus *Lactobacillus* increases significantly in patients with severe COPD [195]. In another paper, Sze and group observed reduced microbial diversity in COPD subjects, which was associated with emphysematous

destruction, remodeling of the bronchiolar and alveolar tissue, and immune infiltration. In line with previous studies, they also found that there was a relative expansion of *Proteobacteria* and *Actinobacteria* phyla, in contrast to the diminishing *Firmicutes* and *Bacteroidetes* phyla in COPD patients when compared to controls. Furthermore, they identified 10 microbial features, including *Haemophilus influenzae*, that could discriminate between control and very severe COPD lung tissue [196].

The microbiome of the upper and lower respiratory tracts in PLWH and HIV-uninfected individuals have been investigated in several studies [87] [112]. In examining the upper respiratory tract, Li *et al* identified that the composition of the oral microbiome was changed in HIV-infected subjects compared to the uninfected subjects, both before and after ART. They observed that PLWH had higher levels of certain microbes including *Streptococci* and *Lactobacilli* in saliva compared to HIV-negative subjects. They also proposed that ART may be able to directly or indirectly reverse these changes to the salivary microbiome, allowing reconstitution of the oral microbiota [119]. In a study by Dang *et al*, tongue scrapings were sampled in people with untreated HIV infection. They observed higher proportions of pathogenic *Veillonella*, *Prevotella*, *Megasphaera*, and *Campylobacter* species, and reduced commensal *Streptococcus* and *Veillonella* species in PLWH relative to healthy controls [39]. A Lung HIV Microbiome Project study by Beck *et al* used oral washes and BALs to compare the oral and lung microbiomes respectively, in PLWH with and without ART, and HIV-uninfected subjects. They saw that the oral microbiome differed based on HIV status, whereas lung communities sampled by BAL were similar between the groups. Additionally, they found *Streptococcus* and *Actinomyces* to be more prevalent in the ART-naïve PLWH, while Rothia to be more abundant in the ART-treated PLWH [14].

On examining the lower respiratory tract, Lozupone *et al* identified an increased colonization of the lungs with *Tropheryma whipplei* in BAL of PLWH compared to the HIV-uninfected population, and that the relative abundance of *T. whipplei* significantly declined with ART [122]. Xu *et al* observed a decrease in microbial diversity in the HIV small airway epithelium, and this was correlated with reduced FEV1/FVC, but only in HIV-negative individuals. They also observed that PLWH had increased Proteobacteria levels and decreased Bacteroidetes and Firmicutes levels compared with HIV-negative controls [214]. Twigg *et al* examined the BAL of subjects infected with HIV with advanced disease and uninfected controls. In contrast to the findings of Beck *et al*, this group demonstrated decreased alpha diversity (measure of richness and evenness) and greater beta diversity in the BAL of the HIV-infected population. They also observed increased levels of *Prevotella* and *Veillonella* in HIV BAL after 1 year of ART [207]. Other groups have shown that the increased abundance of fungal species such as *Pneumocystis* in the HIV lung can cause airway inflammation and pulmonary function decline [149][143][48]. Together, these reports suggest that HIV infection alters the lung microbiome. Antibiotic exposures, cycles of inflammation and immunosuppression, ART, and repeated pulmonary infections may be

other important predictors of microbial dysbiosis in the HIV lung.

2.2 Overarching Hypothesis

From review of literature, we know that integrative studies of multi-dimensional high throughput "-omics" measurements in the context of HIV-associated COPD are limited. We hypothesize that microbial dysbiosis increases COPD susceptibility in people living with HIV.

2.3 Aim 1

Analyze the airway microbiome of subjects with COPD and/or HIV, with further characterization of alpha and beta diversities.

2.4 Methods

2.4.1 Study Population and Design

Cytological brushings of airway epithelial cells were obtained during bronchoscopy from 76 (18 COPD+HIV+, 16 COPD-HIV+, 22 COPD+HIV- and 20 COPD-HIV-) patients at St. Paul's Hospital, Vancouver, BC. Participants could either volunteer for a research bronchoscopy or be enrolled while undergoing bronchoscopy for various clinical indications (the most common being small pulmonary nodules and idiopathic chronic cough). Bronchoscopy was performed under conscious sedation, where 6 to 8 bronchial epithelial brushings were obtained from the right or left upper lobe in the small airways (diameter $\leq$ 2mm), avoiding areas of disease (i.e. pulmonary nodules). Bronchial epithelial brushings were preserved using DNA/RNA shield from Zymo Research (CA, USA) for preservation of genetic material. All patients underwent spirometry testing according to American Thoracic Society/European Respiratory Society guidelines. This study took place between 2015 and 2019 at the University of British Columbia Centre for Heart Lung Innovation (HLI) at St. Paul's Hospital and is approved by the Providence Health Care Research Institute ethics committee (Certificate H15-02166). COPD was defined as either a respirologist diagnosis of COPD or a pre-bronchodilator forced expiratory volume in 1s (FEV1)/forced vital capacity (FVC) ratio $< 70\%$ at screening, and a smoking history of greater than 10 pack-years without an alternative diagnosis to explain airflow obstruction. PLWH were defined as subjects with documented positive HIV-1 infection.

2.4.2 Microbiome Sequencing and Analysis

Microbiome amplicon sequence variants (ASVs), CpG (region of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the 5' → 3' direction) methylation (Chapter 3), and gene expression (Chapter 3) were measured on the same 76 subjects. Further steps were carried out in triplicate for the independent analysis of (1) the COPD effect by comparing COPD- vs COPD+ subjects, (2) the HIV effect by comparing HIV- vs HIV+ subjects, and (3) the combined COPD & HIV effect by comparing the 4 groups - COPD+HIV+, COPD-HIV+, COPD+HIV- and COPD-HIV-.

The microbiome profiles of airway epithelial cells, quantified using the V4 hypervariable region of the 16s rRNA gene, were obtained using touchdown droplet digital polymerase chain reaction (ddPCR), followed by 16s amplicon sequencing using the Illumina Miseq® platform. Sequencing data (fastq files) was analyzed using QIIME 2™ (https://qiime2.org/.1), a microbiome analysis package that is used for taxonomy classification and translating raw sequence data into useful statistical results. Barcode sequences were assigned to their associated samples (demultiplexed), and Divisive Amplicon Denoising Algorithm (DADA2) was used to correct for Illumina amplicon sequence data, and remove low quality sequence regions. During this step, the sequencing reads were merged and resolved into amplicon sequence variants (ASVs).

Further filtering was done to remove ASVs observed consistently in the PCR controls, and other contaminants such as host mitochondrial or chloroplast sequences, ASVs with significantly fewer sequences than the majority, and ASVs present only in one sample (singletons). Following quality filtering steps, sequences were mapped to taxonomy using a pre-trained naïve Bayes classifier artifact trained against Greengenes (13_8 revision) trimmed to contain only the V4 hypervariable region and pre-clustered at 99% sequence identity. Using the MAFFT program in the QIIME 2™ pipeline, a rooted phylogenetic tree was generated and was consecutively used as input to compute different phylogenetic diversity measures. The final outputs from QIIME 2™ (feature table, taxonomy file and phylogenetic tree) were exported for further analysis in R. Alpha diversity differences between (a) the COPD+ and COPD- group, (b) HIV+ and HIV- group, and (c) the COPD+HIV+, COPD-HIV+, COPD+HIV- and COPD-HIV- groups were measured using the Shannon diversity index (a metric of both community richness and evenness), and Faith's phylogenetic diversity (a measure of biodiversity that incorporates phylogenetic difference between species), and visualized using box plots. Beta diversity was measured using the Bray-Curtis dissimilarity index (a metric of differences in richness between two communities), tested with permutational multivariate analysis of variance (PERMANOVA), and visualized using principal components analysis (PCA).

Average relative taxon abundance comparisons were performed between (a) the COPD+

and COPD- group, (b) HIV+ and HIV- group, and (c) the COPD+HIV+, COPD-HIV+, COPD+HIV- and COPD-HIV- groups at the phylum and genus levels. In the COPD-effect and HIV-effect analyses, P-values were determined using the Mann-Whitney U test, and the Benjamini-Hochberg procedure (false discovery rate method correction) was applied to obtain adjusted P values for multiple comparisons between groups. For the combined COPD & HIV analysis, the Kruskal-Wallis test was used to know if there is a significant difference between groups, followed by the Dunn's test to identify which groups are different. Significant taxon differences were identified at adjusted P value < 0.05.

Next, differentially abundant taxonomic features between COPD, HIV and combined COPD & HIV groups were obtained using Linear discriminant analysis (LDA) Effect Size (LEfSe) (LDA effect size = 2). The LEfSe algorithm uses the non-parametric Kruskal-Wallis and Wilcoxon-rank-sum statistical tests to identify features with significant differential abundance between the groups of interest, ranking them according to the effect size (default LDA threshold = 2). For this analysis, the ASVs were collapsed at the *genus* level based on taxonomy names, then the features most likely to explain differences between the different groups were determined.

The use of culture-independent techniques has led to conflicting opinions about the existence of a distinct lower respiratory tract microbiome. There are studies, however limited, in contention that bacteria in the lungs represent microaspiration of oral microbiota, and/or upper respiratory contamination of lower respiratory tract samples caused by passing a bronchoscope through the oral cavity to obtain lung samples [28] [56] [86][136]. To address these concerns, oral washes were collected before bronchoscopy, in addition to other control specimens such as extraction negatives (which did not contain any samples but only the reagents that were used for the DNA extraction), no-template controls (which contained the reagents used for PCR reaction; the DNA sample was replaced with Ultrapure water), cytolyt controls, bronchoscope channel washes obtained prior to bronchoscopy, and unused cytologic brush controls (brush water controls).

It is important to note here that the above-mentioned specimens, with the exception of no-template controls, were collected only for PLWH. A supplemental analysis of these different sample types was performed to examine the presence of contaminating DNA that was not truly present in the brush samples. In addition, the Decontam R package [40] (using as input the DNA concentrations measured in AEC brush samples and other specimens, which consisted of extraction negatives, bronchoscope channel washes, oral wash controls, cytolyt controls and bronchoscope brushing controls) was used to identify ASVs that were potential contaminants (included in the Appendix).

2.5 Results

2.5.1 Description of the Study Cohort

Table (2.1) is a summary of demographics in this study. The final dataset included samples from 76 subjects (18 COPD+HIV+, 16 COPD-HIV+, 22 COPD+HIV- and 20 COPD-HIV-). Overall, the mean age of subjects was 61.3 ± 11.6 years, with 63.2% male subjects and mean FEV1 of 79.88 ± 19.00% predicted values. Based on their smoking status, the subjects were grouped into current (24 ± 31.6), former(31 ± 40.8) and never (18 ± 23.7) smokers. 2.6% were crack cocaine users, 18.4% were marijuana users, and 1.3% were crystal methamphetamine users. 21.1% of total subjects were on prescription inhaled corticosteroids, 17.1% on long-acting muscarinic antagonists (LAMAs) and long-acting beta-agonists (LABAs), and 26.3% on short-acting beta-agonists (SABAs). As measured in PLWH (n=34), the average CD4 count was 435.29 ± 282.55, 76.47% (26 of 34 subjects) had an undetectable HIV viral load, and 70.6% (24 of 34 subjects) were on antiretroviral therapy.

		COPD+HIV+ (n=18)	COPD-HIV+ (n=16)	COPD+HIV- (n=22)	COPD-HIV- (n=20)
Sex	**Male**	14(77.8)	14(87.5)	12(54.5)	8(40.0)
	Female	4(22.2)	2(12.5)	10(45.5)	12(60.0)
Age(years)∗		58.11(9.50)	56.00(10.11)	67.95(7.17)	60.70(14.82)
BMI∗		23.08 (4.76)	27.12 (2.91)	26.43 (5.96)	25.89 (5.41)
Smoking Status	**Current**	11(61.1)	3(18.8)	9(40.9)	1(5.0)
	Former	5(27.8)	8(50.0)	12(54.5)	7(35.0)
	Never	1(5.6)	4(25.0)	1(4.5)	12(60.0)
Pre FEV1 (% predicted)∗		76.07(23.38)	87.98(15.85)	69.25(17.31)	88.47(14.23)
Pre FEV1/ FVC (%)∗		64.30 (12.78)	76.41 (5.35)	62.53 (9.54)	76.64 (6.19)
CD4 Count∗		437.78 (239.94)	432.50(332.21)		
Undetectable HIV Viral Load	**Yes**	16 (88.9)	10 (62.5)		

Table 2.1: Demographics and Clinical Features. ∗Continuous data expressed as Mean(SD). Categorical data expressed as number(% of column totals). Definition of abbreviations: BMI - Body Mass Index; FEV1 - forced expiratory volume in one second; FVC - forced vital capacity; "Pre" refers to spirometry tests before bronchodilator use.

2.5.2 Results from 16s rRNA Quantification

Fig. 2.1 provides the 16s rRNA gene copies/ng of specimen for the 76 brush samples included in the final dataset. There were no significant differences in 16s rRNA levels between the COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- groups (overall Kruskal-Wallis $P = 0.3$).

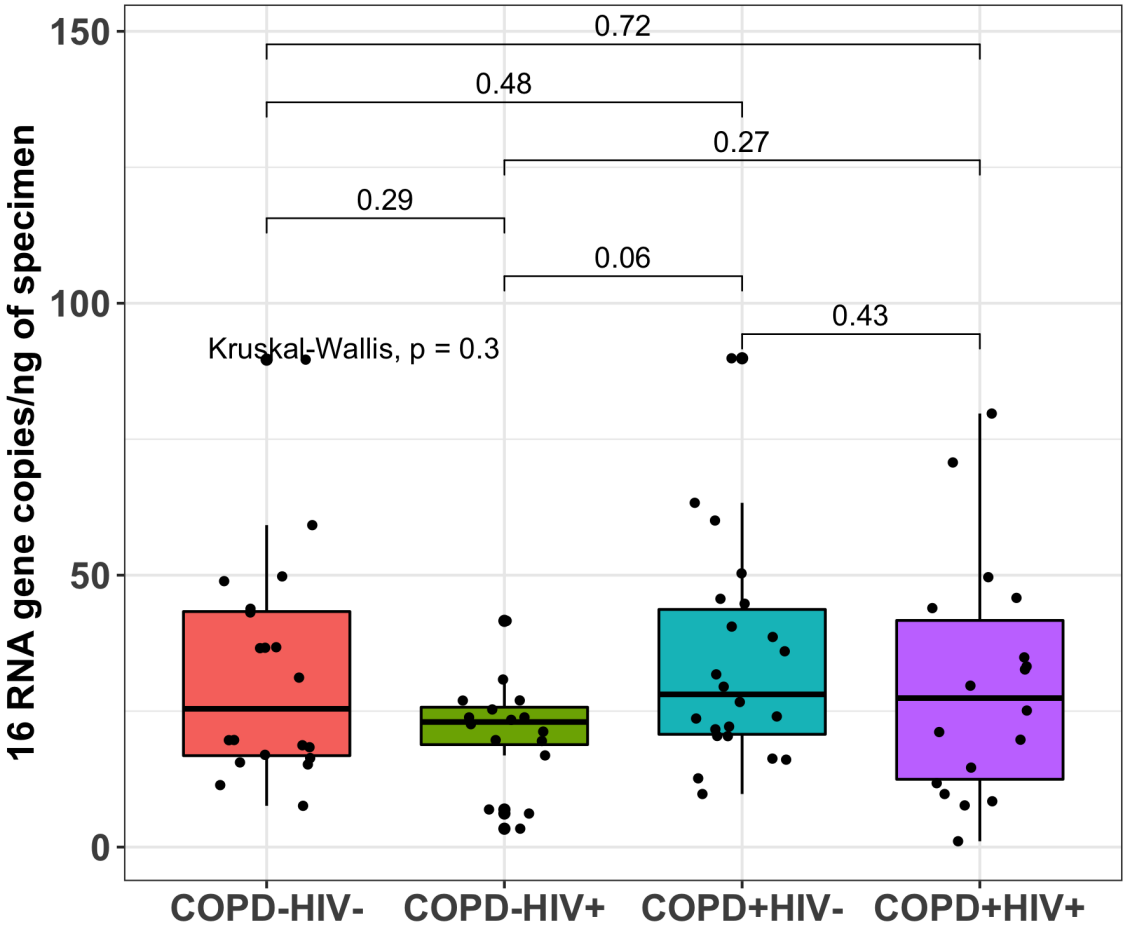


Figure 2.1: 16s RNA gene copies/µL measured in airway epithelial cells. Number of 16s RNA gene copies/µL observed in AEC brush samples grouped by their combined COPD+HIV status. Definition of abbreviations: AEC – Airway Epithelial Cells.

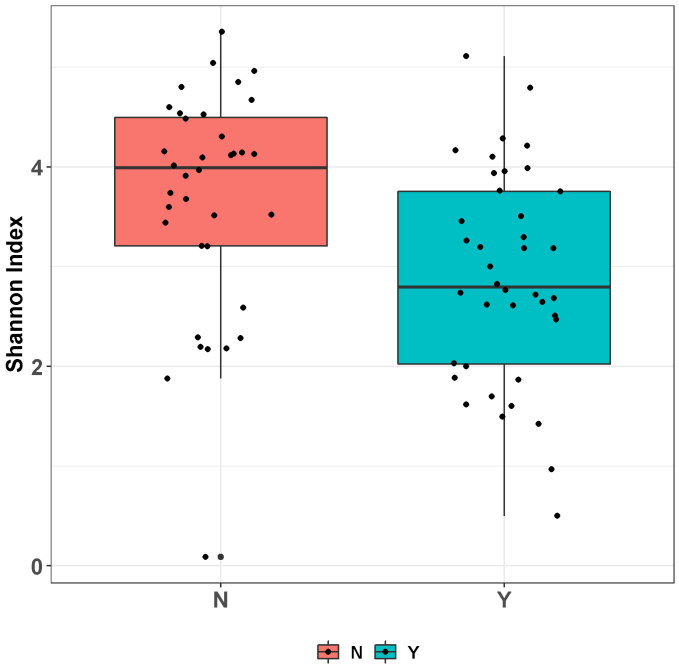
2.5.3 Results from QIIME 2™

Based on the 76 brush samples input into the QIIME 2™ pipeline, the total number of merged reads obtained was 2,005,148 and, after filtering steps (removal of host mitochondrial or chloroplast sequences, ASVs with abundance frequency < 10, singletons, ASVs observed in the controls, and ASVs with no taxonomic annotation at the phylum level) 1,402,568 reads (376 different ASVs) were considered for analysis.

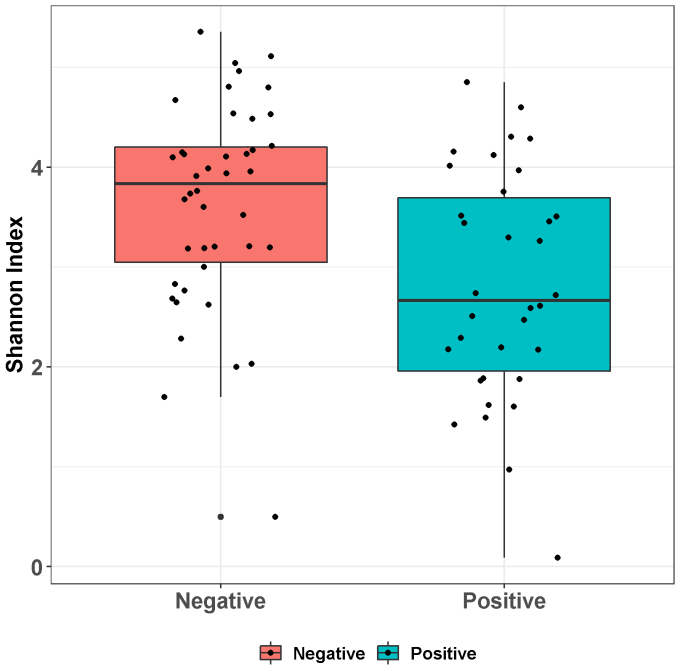
2.5.4 Diversity Metrics

Alpha diversity analysis, measured using the Shannon diversity index as shown in Fig. 2.2, revealed significant differences (expressed as median [interquartile range (IQR)]; Kruskal-Wallis *P-value*) between the COPD+ and COPD- (2.79[1.73] vs. 3.99[1.28]; $P = 0.0013$) groups, HIV+ and HIV- (2.66[1.73] vs. 3.83[1.15]; $P = 0.0023$) groups, and the combined COPD & HIV groups: COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- (2.55[1.60] vs. 3.18[1.32] vs. 3.47[1.93] vs. 4.12[0.91]; $P = 0.0002$). In all three analyses, the groups associated with disease (COPD+, HIV+ and COPD+HIV+ groups) were found to have lower alpha diversity.

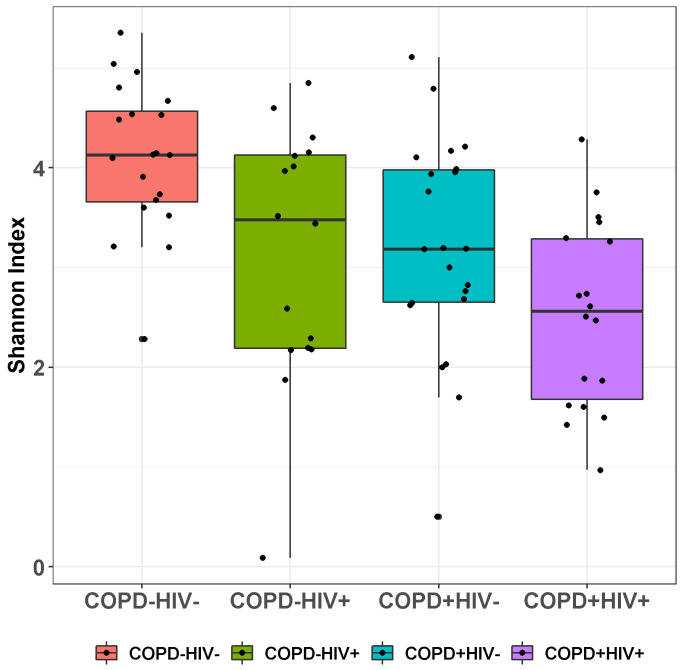
Similarly, Faith's phylogenetic diversity, as shown in Fig. 2.3, also revealed significant differences (expressed as median [interquartile range (IQR)]; Kruskal-Wallis *P-value*) between the COPD+ and COPD- (2.58[1.73] vs. 3.99[2.31]; $P = 0.0004$) groups, HIV+ and HIV- (2.56[1.93] vs. 3.87[2.31]; $P = 0.0036$) groups, and the combined COPD & HIV groups: COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- (2.13[0.98] vs. 2.82[2.05] vs. 3.60[1.85] vs. 4.72[1.99]; $P = 0.0001$). Pairwise Kruskal-Wallis results are in Table (A.1) in the Appendix.


Shannon Index
4
2
0
N
Y
N
Y

(a)

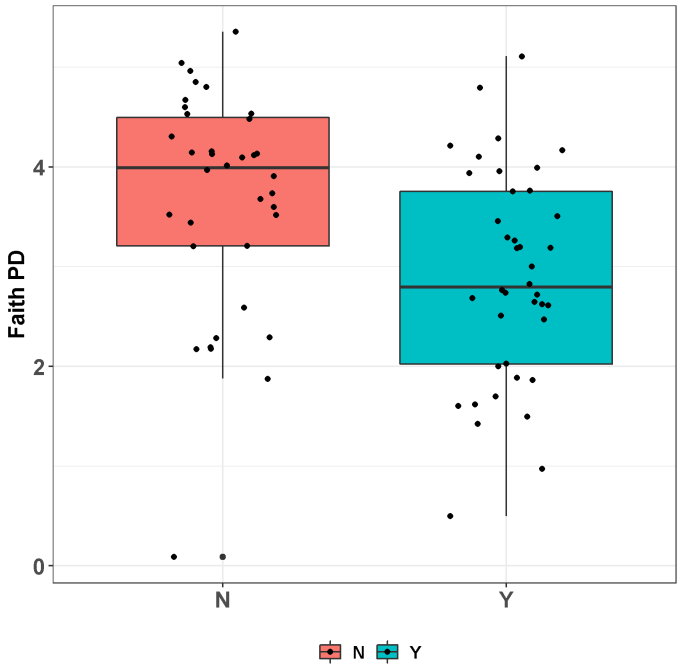

Shannon Index
4
2
0
Negative
Positive
Negative
Positive

(b)

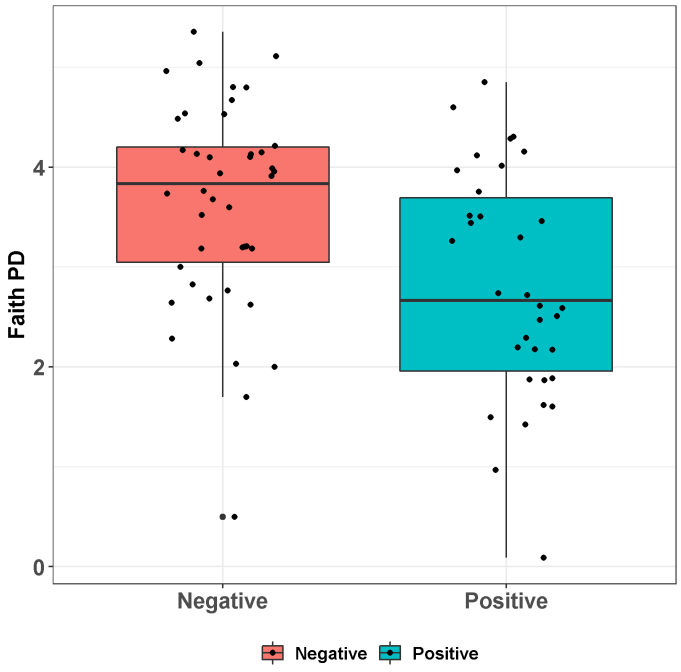


(c)

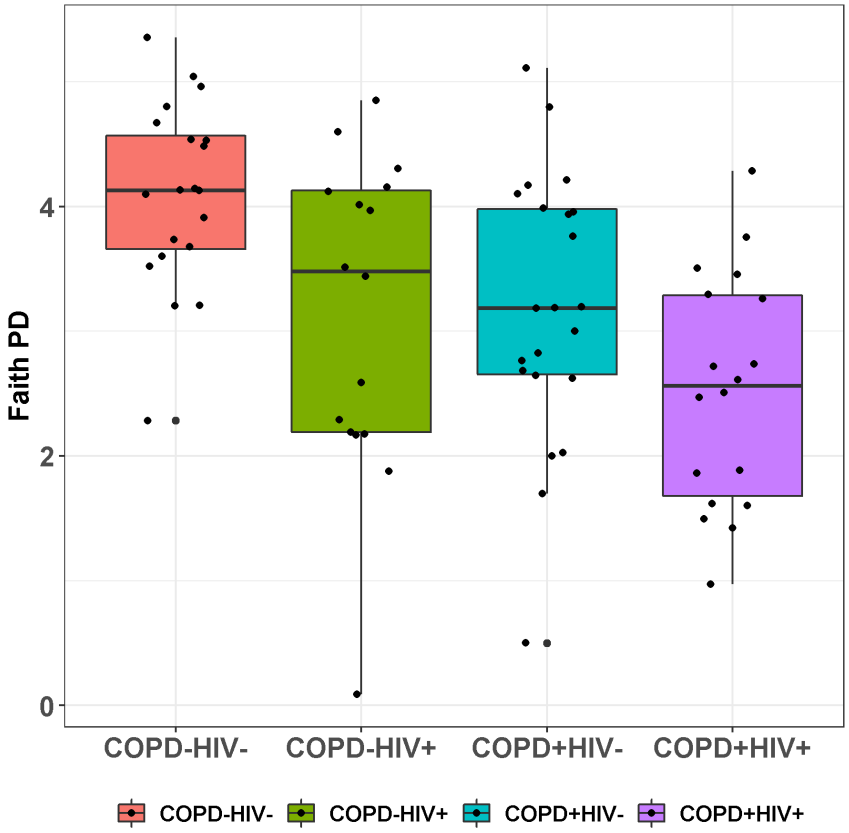
Figure 2.2: Alpha diversity measured using Shannon Diversity Index Alpha diversity differences between (a) COPD+ and COPD- groups ($P = 0.0013$), (b) HIV+ and HIV- groups ($P = 0.0023$), and (c) COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- groups ($P = 0.0002$).


Faith PD
4
2
0
N
Y
N
Y

(a)


Faith PD
4
2
0
Negative
Positive
Negative
Positive

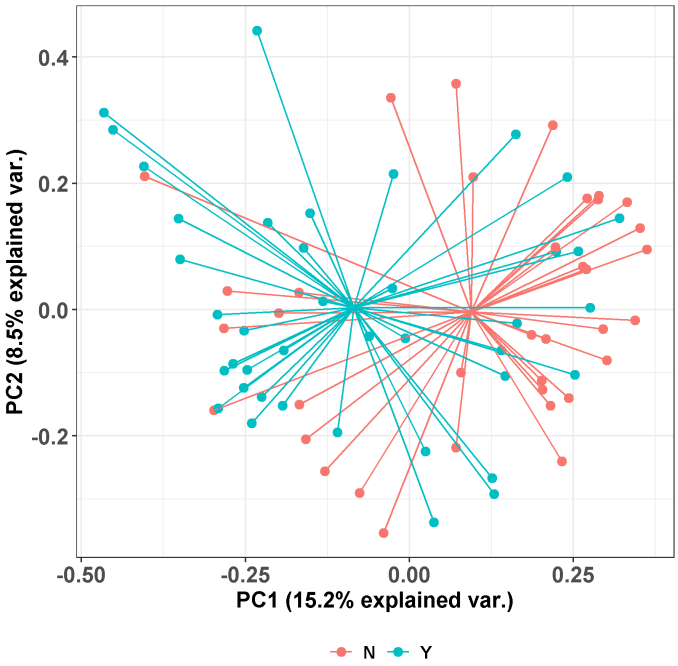
(b)



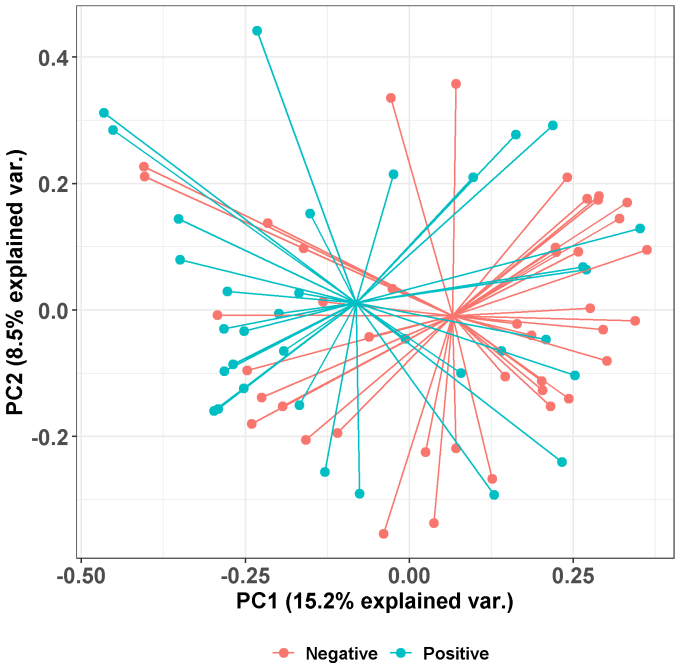
(c)

Figure 2.3: Alpha diversity measured using Faith Phylogenetic Diversity. Alpha diversity differences between (a) COPD+ and COPD- groups ($P = 0.004$), (b) HIV+ and HIV- groups ($P = 0.0036$), and (c) COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- groups ($P = 0.0001$).

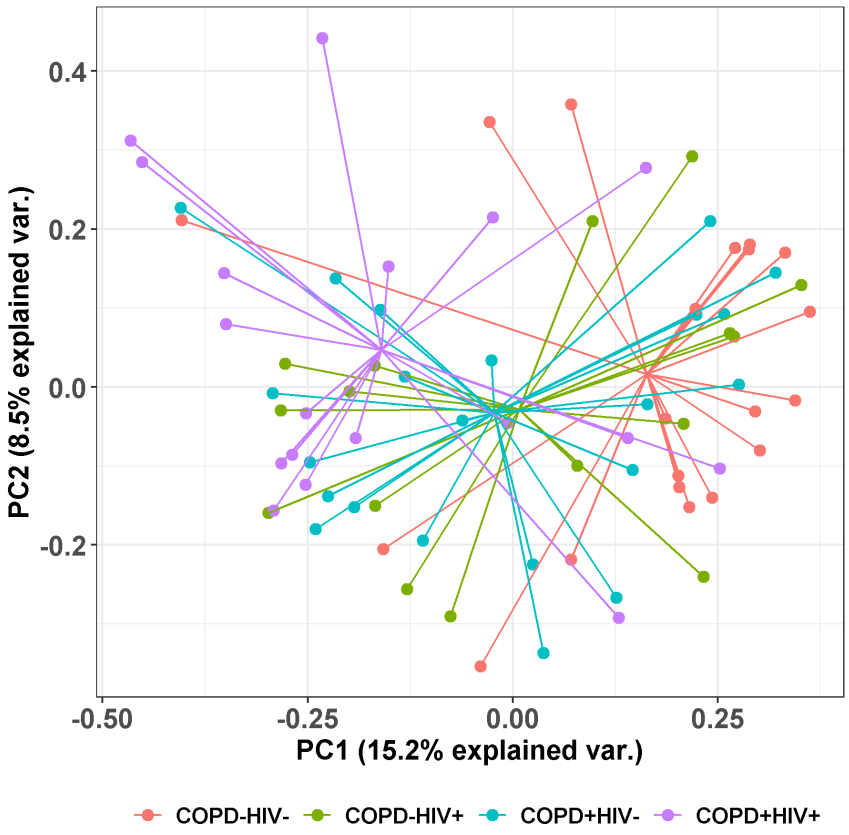
Beta diversity, measured using Bray-Curtis metrics, was used to evaluate the differences in microbial communities between groups. As shown in the principal component analysis plots in Fig. 2.4 and PERMANOVA analysis, there were significant differences between microbiome structures of the COPD+ and COPD- groups ($P = 0.001$), HIV+ and HIV- groups ($P = 0.007$), and the combined COPD & HIV groups ($p = 0.001$). Pairwise PERMANOVA results are included in Table (A.2) in the Appendix.


0.4
0.2
0.0
-0.2
PC2 (8.5% explained var.)
-0.50
-0.25
0.00
0.25
PC1 (15.2% explained var.)
N
Y

(a)


0.4
0.2
0.0
-0.2
PC2 (8.5% explained var.)
-0.50
-0.25
0.00
0.25
PC1 (15.2% explained var.)
Negative
Positive

(b)



(c)

Figure 2.4: Principal component plot showing microbial community structures among subjects based on the Bray-Curtis metric. Microbial community structures in AECs according to (a) COPD status (COPD- subjects (N) – red points; COPD+ subjects (Y) – blue points), (b) HIV status (HIV- subjects (Negative) – red points; HIV+ subjects (Positive) – blue points), and (c) combined COPD & HIV status (COPD+HIV+ – purple points; COPD+HIV- – blue points; COPD-HIV+ – green points; COPD-HIV- – red points) based on Bray-Curtis distances; the centroids for each group are also shown. Definition of abbreviations: PC - principal component.

2.5.5 Taxa abundance

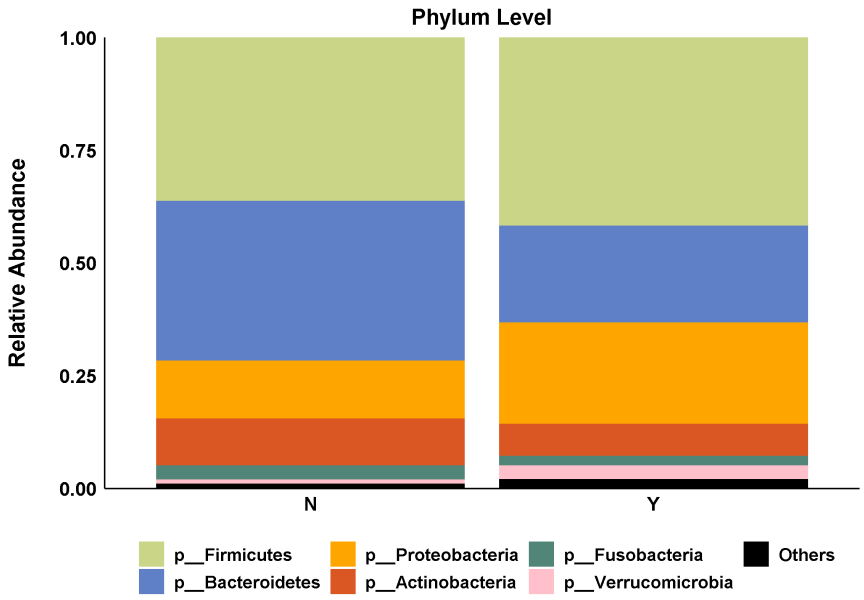
Tables (A.5) and (A.6) in the Appendix show the relative abundance of most abundant phyla and genera in AECs. *Firmicutes* was found to be the most abundant phyla in AECs, followed by *Bacteroidetes*, *Proteobacteria* and others in decreasing order. Similarly, at the genus level, *Prevotella* [f-*Prevotellaceae]*, *Veillonella* and *Streptococcus* were identified as the top three most abundant genera, followed by others.

Relative taxa abundance comparisons at the phylum level between the COPD- and COPD+ groups showed higher relative abundance of *Bacteroidetes* and *Fusobacteria* in the COPD- group (Fig.2.5.a. and Table A.3 in the Appendix). At the genus level, there was a higher relative abundance of *Prevotella* [f-*Prevotellaceae]*, *Veillonella*, *Megasphaera*, *Prevotella* [f-*Paraprevotellaceae]*, *Neisseria*, *Selenomonas* and *Fusobacterium* in the COPD- group, and conversely, a higher relative abundance of *Streptococcus* and *Paenibacillus* in the COPD+ group (Fig.2.6.a. and Table A.4 in the Appendix).

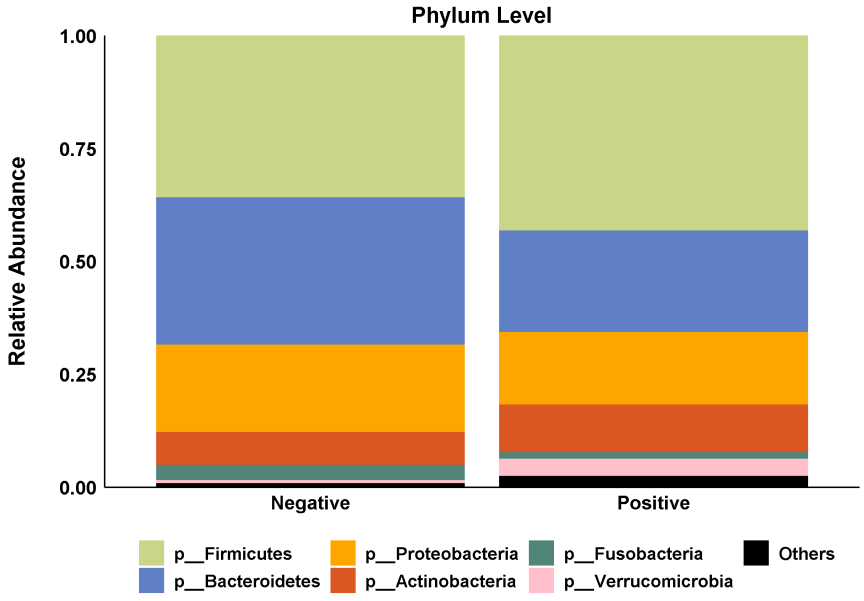
Between the HIV- and HIV+ groups, higher relative abundance of phyla *Fusobacteria* (Fig. 2.5.b. and Table A.7 in the Appendix) and genera *Prevotella*[f-*Prevotellaceae*], *Prevotella*[f-*Paraprevotellaceae*], *Neisseria*, *Selenomonas* and *Fusobacterium* were observed in the HIV- group (Fig.2.6.b. and Table A.8 in the Appendix).

When comparing the COPD-HIV-, COPD-HIV+. COPD+HIV- and COPD+HIV+ groups, phyla *Fusobacteria* and *Bacteroidetes* (Fig. 2.5.c. and Table A.9 in the Appendix), and genera *Prevotella*[f-*Prevotellaceae*] , *Megasphaera*, *Prevotella*[f-*Paraprevotellaceae*], *Neisseria*, *Selenomonas* and *Fusobacterium* showed a significant differences between the 4 groups, with a higher relative abundance in the COPD-HIV- group (Fig.2.6.c. and Table A.10 in the Appendix).

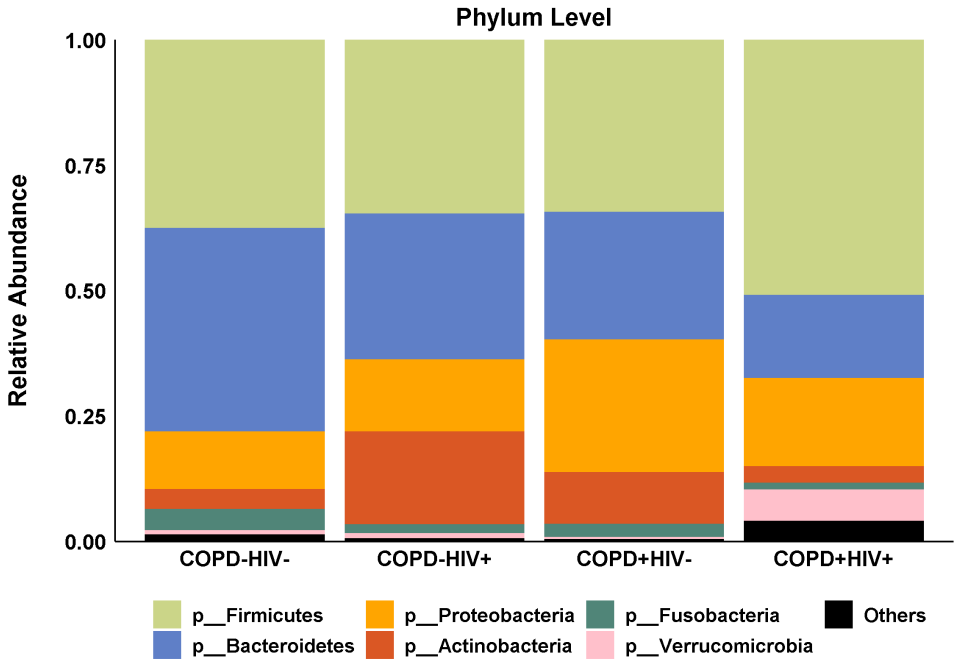
Further pairwise comparisons between groups with corrections for multiple testing (displayed as Dunn's test adjusted *P-values*) showed that phyla *Fusobacteria* and *Bacteroidetes*, and genera *Prevotella*[f-*Paraprevotellaceae*], *Prevotella*[f-*Prevotellaceae*], *Fusobacterium*, *Selenomonas*, *Neisseria* and *Megasphaera* significantly differed between the COPD-HIV- and COPD+HIV+ groups. Other significant genera include - *Prevotella*[f-*Prevotellaceae*] (between the COPD-HIV- and COPD+HIV- groups), *Prevotella*[f-*Paraprevotellaceae*] (between the COPD-HIV- and COPD-HIV+ groups), and *Paenibacillus* (between the COPD-HIV+ and COPD+HIV+ groups) (Tables A.11) and Table A.12 in the Appendix).


Phylum Level
Relative Abundance
1.00
0.75
0.50
0.25
0.00
N
Y
p__Firmicutes
p__Bacteroidetes
p__Proteobacteria
p__Actinobacteria
p__Fusobacteria
p__Verrucomicrobia
Others

(a)

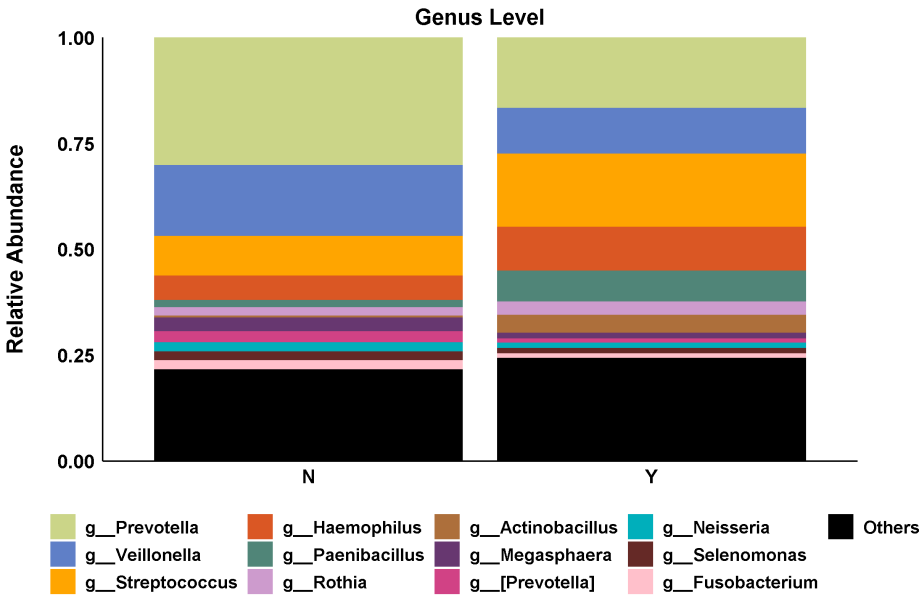

Phylum Level
Relative Abundance
1.00
0.75
0.50
0.25
0.00
Negative
Positive
p__Firmicutes
p__Bacteroidetes
p__Proteobacteria
p__Actinobacteria
p__Fusobacteria
p__Verrucomicrobia
Others

(b)

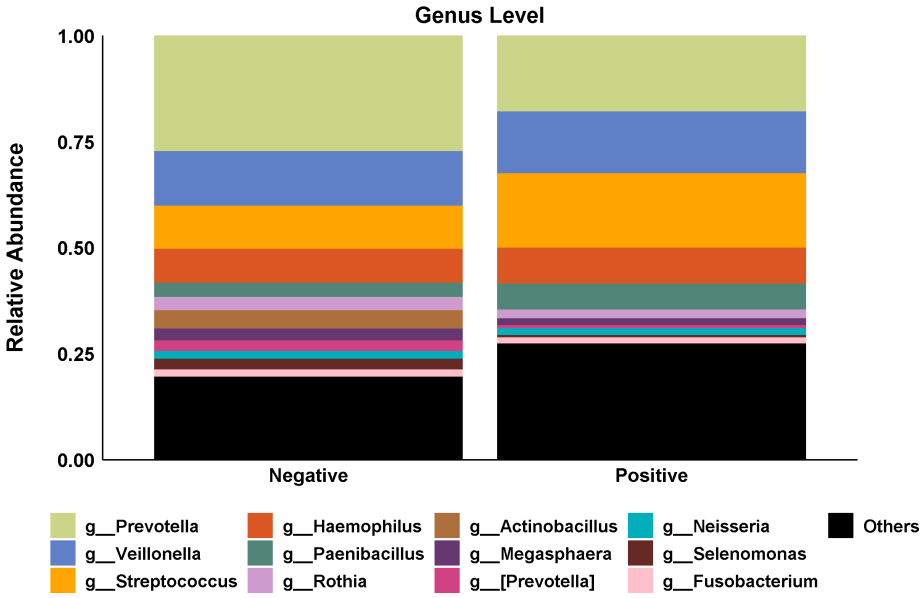


(c)

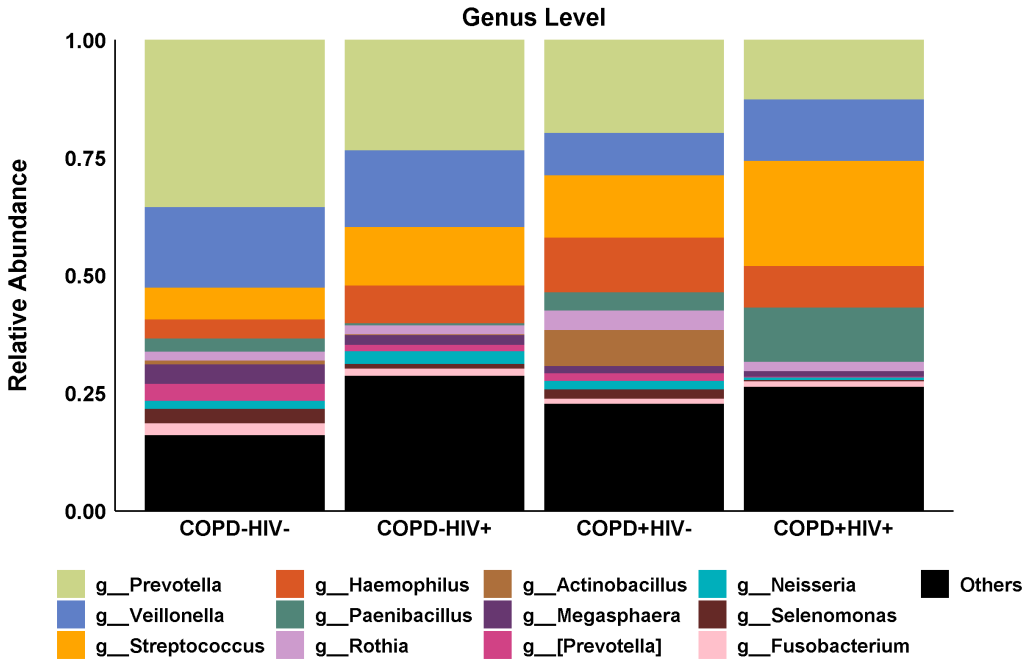
Figure 2.5: Average relative taxa abundance comparisons between (a) COPD+ and COPD- groups, (b) HIV+ and HIV- groups, and (c) COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- groups at the phylum level.


Genus Level
Relative Abundance
1.00
0.75
0.50
0.25
0.00
N
Y
g__Prevotella
g__Veillonella
g__Streptococcus
g__Haemophilus
g__Paenibacillus
g__Rothia
g__Actinobacillus
g__Megasphaera
g__[Prevotella]
g__Neisseria
g__Selenomonas
g__Fusobacterium
Others

(a)


Genus Level
Relative Abundance
1.00
0.75
0.50
0.25
0.00
Negative
Positive
g__Prevotella
g__Veillonella
g__Streptococcus
g__Haemophilus
g__Paenibacillus
g__Rothia
g__Actinobacillus
g__Megasphaera
g__[Prevotella]
g__Neisseria
g__Selenomonas
g__Fusobacterium
Others

(b)

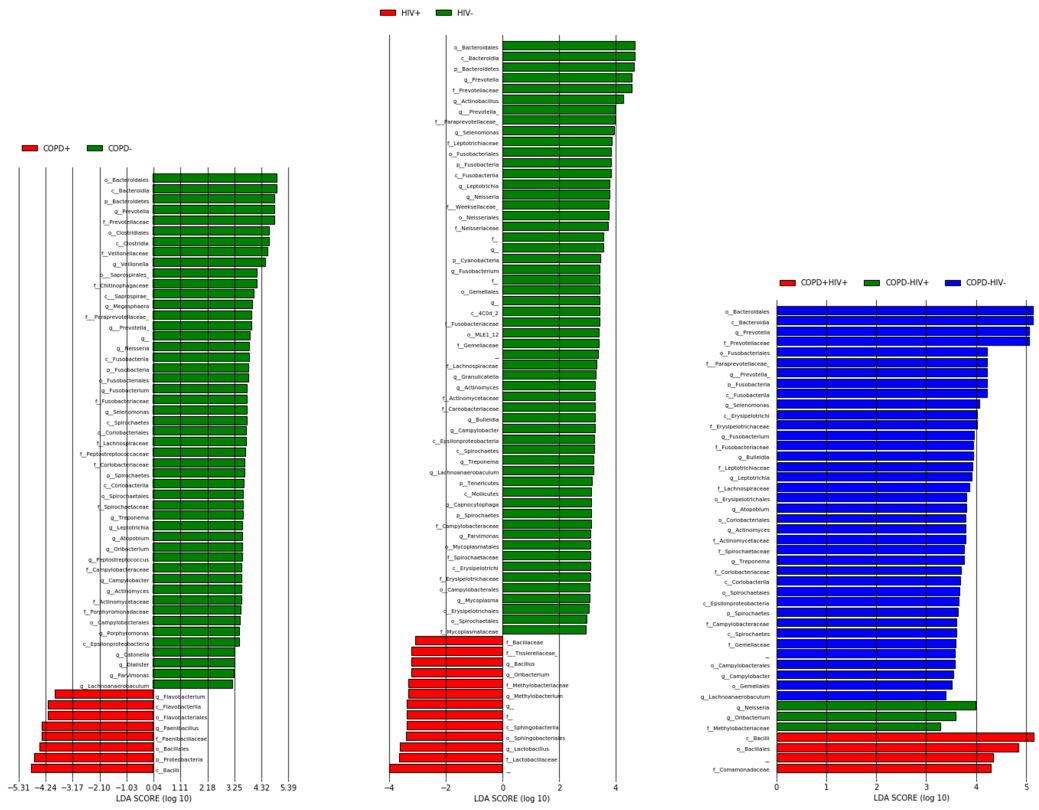


(c)

Figure 2.6: Average relative taxa abundance comparisons between (a) COPD+ and COPD- groups, (b) HIV+ and HIV- groups, and (c) COPD+HIV+, COPD+HIV-, COPD-HIV+ and COPD-HIV- groups at the genus level.

2.5.6 LEfSe

LEfSe analysis was performed on 126 ASVs, obtained after collapsing features at the genus level based on taxonomy. As seen in Fig.2.7, this analysis identified 57, 69 and 45 discriminating taxon features between the COPD groups, HIV groups, and combined COPD & HIV groups, respectively, across different taxonomic levels. No unique features were identified in the COPD+HIV- group.



(a) (b) (c)

Figure 2.7: Differential taxa features identified by LEfSe (LDA effect size = 2) (a) COPD (COPD- group – green bars; COPD+ group – red bars), (b) HIV (HIV- group – green bars; HIV+ group – red bars) and (c) combined COPD & HIV status (COPD+HIV+ group - red bars, COPD-HIV- group - blue bars, and COPD-HIV+ group - green bars); The length of the bar represents a log10 transformed LDA score.

2.6 Supplemental Analyses Comparing Different Specimen Types

In the supplemental analysis of different specimen types obtained from only HIV+ subjects, pairwise comparisons showed that bronchial brushings had significantly different 16s rRNA counts from bronchoscope channel washes (P = 4.1e-11), bronchoscope brushing controls (P = 0.00022), cytolyt controls (P = 4.2e-11) and extraction negative specimens (P = 0.0048); no such difference was found between brushings and oral wash control samples (overall Kruskal-Wallis $P <$ 2.2e-16) (Fig. 2.8). Fig. 2.9 showed the difference of microbial community structures among the different specimen types based on Bray-Curtis distance. Bronchial brushings were significantly different from control specimens in HIV+ subjects (PERMANOVA P = 0.001). Pairwise PERMANOVA comparisons between the different sample types are shown in Table (A.13) in the Appendix.

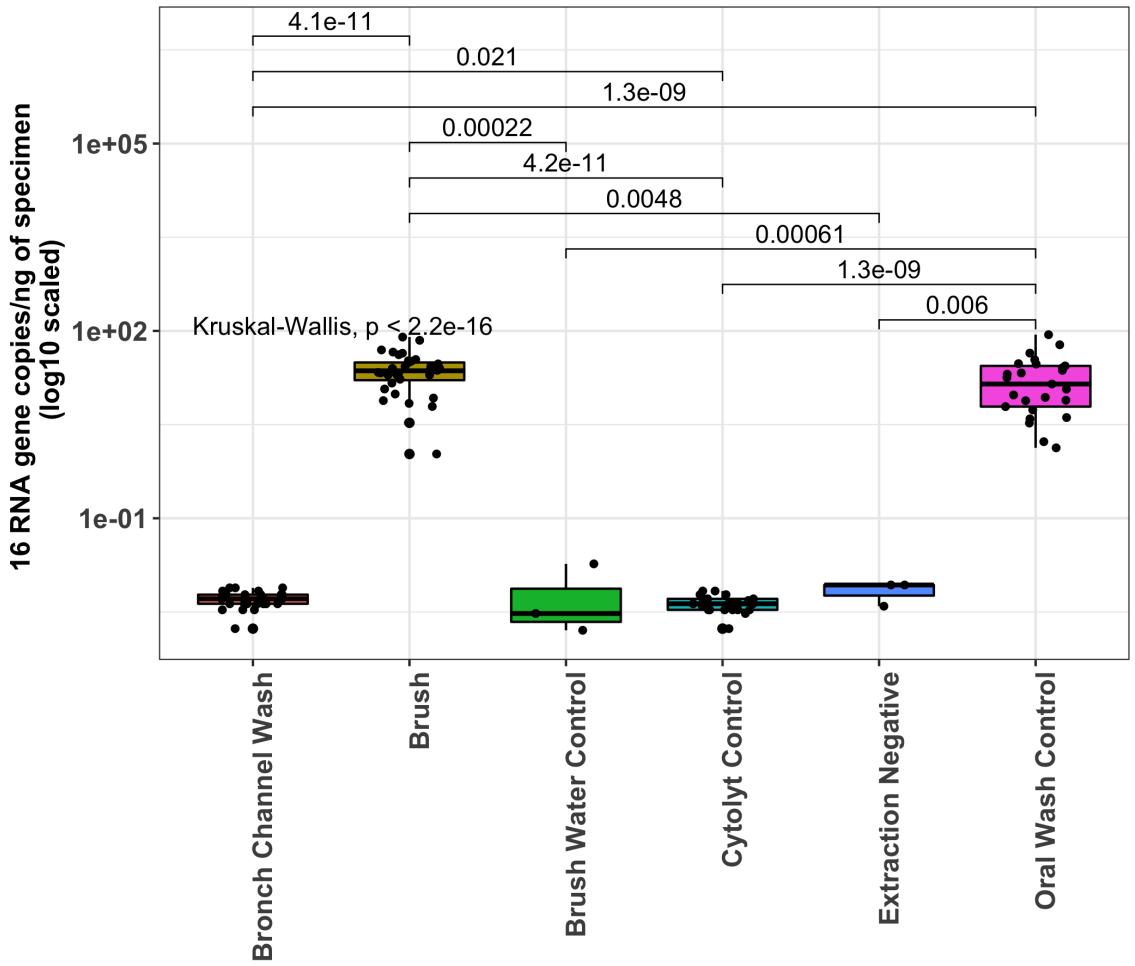


Figure 2.8: 16s RNA gene copies/µL in bronchial brushings and control specimens, consisting of bronchoscope channel washes, brush water controls, cytolyt controls, extraction negatives, and oral wash controls. Only significant Kruskal-Wallis pairwise comparison values are shown.

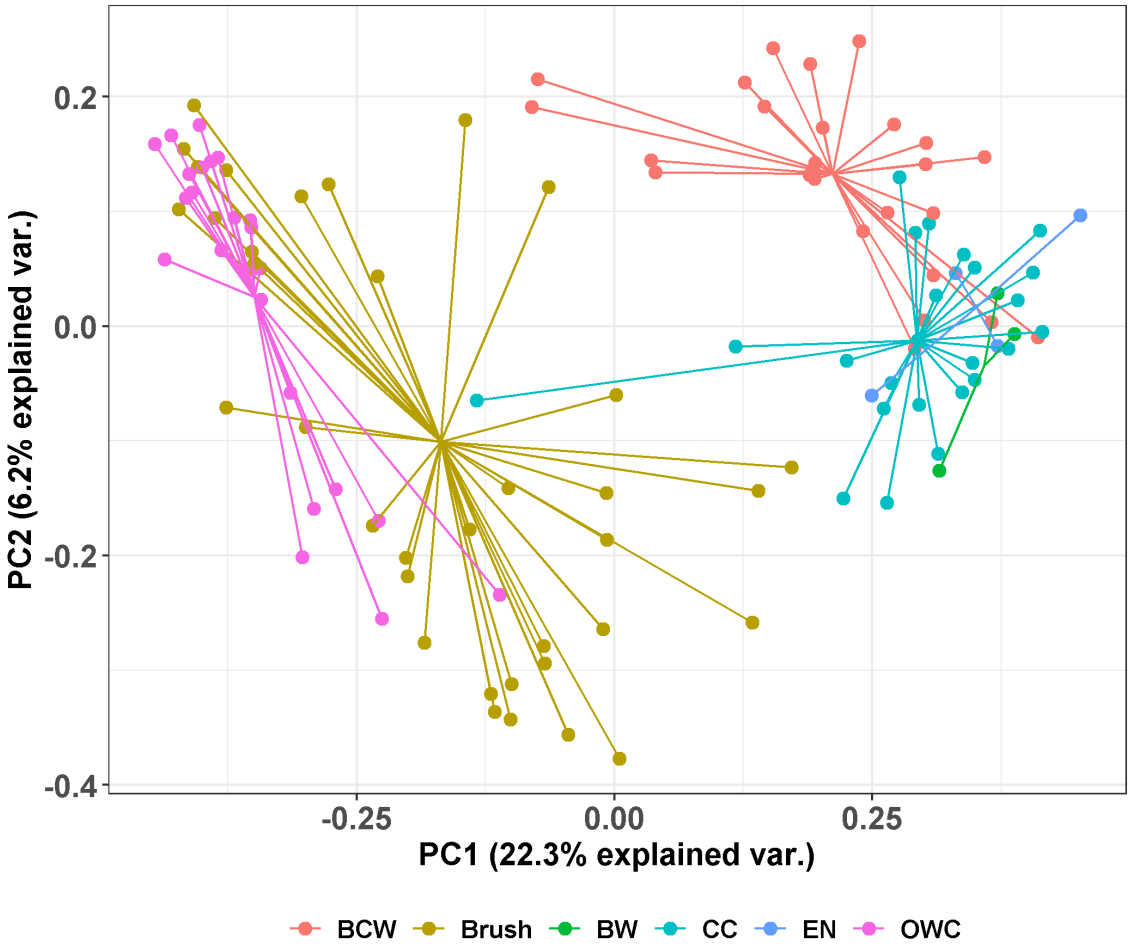


Figure 2.9: Principal component plot showing microbial community structures among specimen types in HIV+ subjects based on the Bray-Curtis metric. The centroids for each group are also shown. Definition of abbreviations: PC - principal component; BCW - bronchoscope channel wash; BW - brush water control; CC - cytolyt control; EN - extraction negative; OWC - oral wash control.

2.7 Discussion

The mechanism of HIV-associated COPD is not fully understood, with many factors - increased risk factor behaviors, apoptosis, microbial colonization, and altered inflammatory and oxidant–antioxidant responses, among others - playing a possible role in disease pathogenesis. Together, these factors alter the lung microbiome, which is critical in performing a wide range of metabolic activities that benefit the host.

From our analysis of the small airway microbiome, we found that the subject groups most associated with disease (COPD+, HIV+ and COPD+HIV+ groups) had reduced average species diversity, and significantly disrupted microbial communities from their relatively "healthy" counterparts (COPD-, HIV- and COPD-HIV- groups). This is in consensus with the idea that composition and diversity are important components of a "healthy" microbiome [121].

Consistent with previous 16s rRNA studies, we also identified that the "healthy" groups were enriched in characteristic phyla *Fusobacteria* (in COPD-, HIV- and COPD-HIV- groups) and *Bacteroidetes* (only in COPD- and COPD-HIV- groups). The decrease in levels of these microbes may be associated with obstructive lung disease in PLWH. Sze *et al*, Xu *et al* and Ramsheh *et al* observed a similar decrease in *Bacteroidetes* levels, in addition to increased *Proteobacteria* and decreased *Firmicutes* levels, in GOLD Stage 4 COPD patients, PLWH, and COPD patients with or without ICS, respectively, when compared with healthy controls [196] [214] [160]. In our analyses however, we did not observe significant differences in abundance levels of *Proteobacteria* and *Firmicutes* between the different groups.

At the genus level, *Prevotella*, *Selenomonas*, *Neisseria* and *Fusobacterium* were more abundant in the "healthy" groups, and these results are similar to those of many previous studies [179] [28] [45] [25] [114] [13]. *Prevotella* is a widely studied genus, however its exact role in the respiratory system still remains unknown. These microbes are enriched in the oral cavity, and their colonization of lower airways most likely results from microaspiration [179]. *Prevotella* has been described both in the context of health and disease. On one hand, *Prevotella* abundance has been described in association with better lung function, and reduced dyspnea scores and inflammation [179] [181] [160]. Many of these properties of *Prevotella* are through its interactions with other members of the microbiota, indicating that dynamics of this genera might be important in regulating inflammation and dysbiosis [109]. On the other hand, certain strains that exhibit pathobiontic properties have been implicated as promoters of subclinical inflammation, particularly Th-17 inflammation [179] [107] [108]. Twigg *et al*, in a small cohort of nine subjects infected with HIV, observed that long-term (3 years) ART use, which would normally be associated with a more "healthier" phenotype, was associated with decreased *Prevotella* abundance [207]. On the contrary, our

findings demonstrate that "disease" groups (COPD and/or HIV subjects) are associated with increased *Prevotella* when compared to the healthy groups. However, it is possible that this observed decrease in *Prevotella* in our cohort could be confounded by duration of ART usage, which we were unable to capture.

Other important genera included *Veillonella* and *Streptococcus*, which have been associated with pro-inflammatory cytokine production and promoting vascular injury. These microbes, along with *Neisseria* were found to be increased in PLWH with abnormal lung function [219]. Furthermore, it has been shown that microaspiration of these anaerobic commensals from the oral cavity seeds the lower airways, and drives clinical presentation of lung injury [11] [140] [185]. Our results echoed these studies for genera *Streptococcus* and *Neisseria*. We however observed higher relative abundance of *Veillonella* in the COPD-group (no significant results in the HIV effect and combined COPD+HIV effect analyses).

Among other major genera in the HIV lung is *Fusobacterium*. Many species of this genera have been described as part of the healthy core lung microbiome, and have been indicated as interconnectors between human and bacterial cells [55]. *Fusobacterium* also supports the growth of anaerobic organisms in oxygenated environments [22]. We hypothesize that this may be crucial in the early stages of disease when the lung is colonized by anaerobic oral commensals. In the aforementioned study by Engel *et al*, *Fusobacterium* was also found to cluster with other genera such as *Megasphaera*, a member of the healthy lung microbiome. This genus can beneficially affect the host by modulating short chain fatty acid production, reducing airway inflammatory responses, and preventing colonization of respiratory pathogens [210].

In summary, the relative abundance of phyla *Fusobacteria* and *Bacteroidetes* change significantly with disease, and these changes were noted down to the level of genera such as *Prevotella*, *Streptococcus*, *Veillonella*, *Selenomonas*, *Neisseria*, and *Fusobacterium*. Thus the depletion of taxa that typically constitute a "healthy" microbiome, with the concurrent establishment of potentially harmful taxa may be correlated with host physiological parameters in disease, and might be responsible for driving HIV-associated COPD disease pathology. It also underscores the potential pathogenicity of oral commensal microbes and their proliferation in the lungs of HIV-infected patients. Overall, alterations in the respiratory microbiome may be at least partially responsible for injury seen in PLWH, although we did not prove causation in this study.

Chapter 3

The Microbiome and Other -*Omes*

3.1 Introduction

Prior work and our results show that significant differences in the respiratory microbiome may be linked to COPD pathogenesis in PLWH. In may also be that microbiome-host interactions play a vital role in determining disease severity and symptoms. In this chapter, we describe the analysis of the host methylome and transcriptome in subjects with COPD and/or HIV.

3.1.1 Methylome

The 'methylome' denotes the nucleic acid methylation modifications in an organism's genome or in a particular cell. DNA methylation is an epigenetic mechanism, where a methyl group is added to the C5 position of cytosine residues in CpG dinucleotide sequences. Regulated by a family of DNA methyltransferase enzymes, DNA methylation contributes to the transcriptional on/off state of genes, and can thereby affect its expression and/or its function [90]. Methylation patterns are heritable, but a growing body of evidence indicates that environmental exposures may influence epigenetic regulation of gene expression, and thereby determine phenotype. [161] [111].

DNA methylation takes place both at the promoters and within the gene body (intragenic). It is well-known that transcriptional silencing is caused by the hypermethylation of promoter regions of CpG islands (CpGIs), which are regions of DNA that contain an increased frequency of CpG dinucleotides (CG content $> 50\%$ and observed/expected CpG ratio $> 60\%$) [6]. This transcriptional repression of the viral promoter is a prevailing regulatory mechanism in the HIV provirus, and is the major cause of HIV-1 latency [29]. A recent study by Kint *et al* conducted on peripheral blood mononuclear cells evaluated the lesser-known role of intragenic methylation, and found a general pattern of low promoter methylation and higher intragenic methylation. In addition, they observed that ART-naïve seroconverters showed increased promoter methylation and decreased intragenic methylation compared with long-term ART-treated individuals, individuals in the chronic phase of the infection, and ART-naïve long-term non-progressors [98]. This data suggests that intragenic DNA methylation could be a late event during infection, and may be involved in

the regulation of intragenic promoters, alternative splicing, and in the activation of retroviruses, repetitive elements, and prevention of aberrant transcript production [146] [92].

Epigenetic patterns in PLWH are distinctly different from the normal population. Leung *et al* observed telomere shortening and methylation changes in peripheral blood during the short period immediately following HIV seroconversion but not between post-HIV seroconversion and a later follow-up time point [117]. This is consistent with other studies also conducted on blood samples, which showed results of accelerated ageing in PLWH [145] [73]. More recently, members of our group identified that ART-naive PLWH with CD4 T-cell counts $> 500/\mu L$ show evidence of advanced methylation ageing. They also found differentially methylated positions (DMPs) that corresponded to genes highly enriched for cancer-related pathways [217].

Only a few studies have examined methylation changes in the HIV lung. A recent report by our group, based on methylation at CpG-rich genomic sequences Alu and LINE-1, found that PLWH with airflow obstruction have greater hypomethylation and accelerated ageing in comparison to those with normal lung function. The identified DMPs were enriched for biological pathways related to chronic viral infections and small airway remodelling in COPD [32]. In examining the relation between the microbiome and methylome, Xu identified distinct microbiome and methylation profiles between the small airway epithelial cells of PLWH and uninfected subjects. The two groups also had different methylation patterns in genes involved in cell differentiation, biological processes such as cell-cell adhesion and cell surface receptor signaling pathways, and cellular components such as microtubule cytoskeleton and nucleoplasm [216]. In a similar work, Yang examined lung tissue samples between the HIV+ and HIV- groups and found that the former group had significantly shorter absolute telomere length (aTL). However, a microbiome analysis conducted concurrently showed that there was no correlation between bacterial load and aTL between the two groups [220].

3.1.2 Transcriptome

The 'transcriptome' refers to the set of all RNA molecules, such as messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and other non-coding RNA molecules that are present in cells [199]. The integrity of gene and subsequent protein expression are primarily maintained by epigenetic processes such as histone modification and DNA methylation–mediated transcriptional regulation, among other kinds of chemical modifications [66]. The crosstalk between these epigenetic processes, especially, can activate or silence gene expression by influencing transcriptional activity and how DNA strands are packaged [135].

HIV infection causes changes in gene expression profiles in the lungs of PLWH. A recent

study by Chung *et al*, conducted on small airway epithelial cells exposed to *in vitro* HIV, showed that HIV upregulates the expression of inflammatory genes in a dose-dependent manner. The subsequently released infammatory mediators IL-8, IL-1β, ICAM-1 and GM-CSF also induced the migration of alveolar macrophages and neutrophils [31]. In evaluating the relationship between the microbiome and transcriptome, Sze *et al* examined the gene expression patterns of the host in response to the bacterial microbiome using bronchial epithelial cells obtained from 21 PLWH. They identified that measures of the airway microbiome including alpha diversity measures, phyla, and OTUs, had strong associations with gene modules enriched for immune and inflammatory responses, cell signaling, and cilia pathways. In particular, they observed that *Firmicutes* and *Proteobacteria* phyla antagonistically regulated gene expression pathways related to oxidation/reduction and intracellular organelles [197]. A recent report by Bhadriraju *et al* looked into the microbe-mediated modulation of epithelial gene expression in sputum of HIV-positive children with clinical indicators of chronic lung disease. They found that children with *Haemophilus*-, *Moraxella*-, or *Neisseria*-dominated sputum microbiota had increased expression of pro-inflammatory cytokines such as IL-1β, IL-33 and E-cadherin, and decreased expression of anti-inflammatory Muc5AC [15].

3.2 Hypothesis

Microbial dysbiosis in the airway epithelium of PLWH is accompanied by changes in host response. We hypothesize that differential DNA methylation and gene expression are associated with heightened COPD susceptibility in PLWH.

3.3 Aim 2

Analyze the airway methylome and transcriptome of subjects with COPD and/or HIV to identify top differentially regulated CpGs and genes, respectively, and evaluate their corresponding biological roles.

3.4 Methods

3.4.1 Study Population and Design

Cytological brushings of airway epithelial cells obtained from the same 76 subjects with COPD and/or HIV were used to quantify gene expression and CpG methylation. Further

steps in the transcriptome and methylome analyses were carried out in triplicate for the independent analyses of - (1) the COPD effect by comparing COPD- vs COPD+ subjects, (2) the HIV effect by comparing HIV- vs HIV+ subjects, and (3) the interactive COPD*HIV effect by comparing the 4 groups : COPD+HIV+, COPD-HIV+, COPD+HIV- and COPD-HIV-.

3.4.2 Methylome Analysis

DNA Methylation Profiling and Data Analysis

DNA was extracted from airway epithelial cells, followed by sodium bisulphite conversion of unmethylated cytosines to uracils. Subsequently, methylation profiles were obtained using the Illumina Infinium Methylation EPIC BeadChip®, which covers over 850,000 CpG sites across the genome [158]. CpG methylation was measured as methylation beta-values (β), defined as the ratio of the methylated probe signal to the overall signal, ranging from 0 (completely unmethylated) to 1 (fully methylated), and logit-transformed into M-values (log2 ratio of the intensities of methylated probe versus unmethylated probe) for statistical analyses [52]. In filtering for quality, probes were removed from downstream analyses based on detection P-values ($<$ 1e-10), chromosomal location (XY chromosome-linked and non-CpG), presence of single nucleotide polymorphisms in the probe sequence, and cross-hybridization [203]. Further processing included normal-exponential out-of-band (noob) background correction of probe intensity data [204], beta-mixture quantile normalization [201], batch effect correction using ComBat (sva 3.30.1) [91], and removal of outlier CpG probes using Gaphunter (R package "minfi" 1.28.4) [4].

Identifying Differentially Methylated Probes and Regions

Covariate selection was performed to identify any confounding variables. No covariates were found to have a significant influence of CpG methylation. To identify significantly differentially methylated probes (DMPs) or CpG sites in (1) COPD- vs COPD+ patients, (2) HIV- vs HIV+ patients, and (3) COPD+HIV+, COPD-HIV+, COPD+HIV- and COPD-HIV- patient groups, respectively, the following robust linear models (not adjusted for any covariates) were used:

(1) Methylation (M value) $\sim$ COPD status + EPISTRUCTURE PC1–PC5
(2) Methylation (M value) $\sim$ HIV status + EPISTRUCTURE PC1–PC5
(3) Methylation (M value) $\sim$ COPD*HIV status + EPISTRUCTURE PC1–PC5

DMPs were identified at a false discovery rate (FDR) $<$ 0.10. DNA methylation changes often occur in contiguous genomic regions called differentially methylated regions (DMRs).

DMRs, defined as ≥ 3 contiguous CpG probes, were identified using DMRcate (R, Bioconductor 1.18.0) at a significance threshold of FDR < 0.10.

Volcano plots displaying unstandardized signal (beta difference) against noise-adjusted/standardized signal (-log(10)(p-value) were used to visualize genes associated with top CpGs that were differentially methylated between the COPD and HIV groups. Candidate genes were applied to annotation databases Kyoto Encyclopedia of Genes and Genomes (KEGG; https://www.genome.jp/kegg/) and Gene Ontology (GO; http://geneontology.org/) to identify significant pathways at an FDR < 0.10.

3.4.3 Transcriptome Analysis

Gene Expression Profiling and Data Analysis

Total RNA was extracted from cytological brushings, and whole-transcriptome sequencing was carried out using the Illumina NovaSeq 6000® RNA sequencing system. Raw sequence data was controlled for quality using FastQC [172]. STAR (Spliced Transcripts Alignment to a Reference), an ultra-fast RNA-seq quasi-alignment software package [47], was used to align the paired end reads to the GRCh37 reference genome in GENCODE (version 31), which contains 20,687 protein-coding and 9640 long non-coding RNA loci [75]. Transcript-level quantifications were aggregated to obtain gene-level counts using the R package *tximport* [192]. Principal component analysis (PCA) was used as a visual tool to evaluate batch effects. Count data was normalized to log2-counts per million (log2-CPM) reads using limma voom [163], and genes with low expression were filtered out, such that log2-CPM is > 1 in at least a fourth of the total samples.

Identifying Differentially Expressed Genes

One variable was found to have a significant effect on gene expression - *Sex*. To identify differential gene expression in the analysis of (i) independent COPD effect (using only HIV- subjects, n = 42), (ii) independent HIV effect (using only COPD- subjects, n = 36), and (iii) the interactive COPD*HIV effect (using all subjects, n = 76), respectively, the following robust linear models were used after adjusting for significant covariate "Sex":

(1) Gene expression $\sim$ COPD status + Sex
(2) Gene expression $\sim$ HIV status + Sex
(3) Gene expression $\sim$ COPD*HIV status + Sex

Volcano plots were used to visualize top genes that were differentially expressed between the COPD and HIV groups. Further, top genes were enriched onto annotation databases

Kyoto Encyclopedia of Genes and Genomes and Gene Ontology, to generate a list of significant pathways at an FDR < 0.10.

3.5 Results

3.5.1 Methylome Analyses

PCA was used to compare distances between the respective groups in the independent COPD, HIV, and COPD*HIV effect analyses (Fig. B.1 in the Appendix). Covariate selection analysis identified no significant confounders (Fig. 3.1). Robust linear regression was applied, and differentially methylated positions (DMPs) and regions (DMRs) were identified at FDR < 0.05: 8,736 DMPs and 1,110 DMRs in the COPD analysis, 61,536 DMPs and 6,417 DMRs in the HIV analysis, and 1,755 DMPs and 211 DMRs in the COPD*HIV analysis.

As seen in Fig. 3.2, volcano plots identified the following to be the genes associated with the three most significant DMPs at FDR < 0.10 : genes *MAP7D1* (MAP7 Domain Containing 1), *SMARCD3* (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily D, Member 3) and *TCF7L2* (Transcription Factor 7 Like 2) between the COPD- and COPD+ (reference) groups, and genes *MUC5B* (Mucin 5B, Oligomeric Mucus/Gel-Forming), *DDX39A* (DExD-Box Helicase 39A) and *CCDC150* (Coiled-coil domain-containing protein 150) between the HIV- and HIV+ (reference) groups. The identified DMPs appear to have significant p-values but very small effect sizes in methylation beta change, indicating a minimal disease effect. The question of whether these very small effect sizes are true methylation differences or mere aberrations within the error range of the methylation array raises the possibility that they could imply false signals.

Top CpG-associated genes that were significant in the COPD*HIV analyses FDR <0.10 are in Table. 3.1. Of the 5 CpG sites, COPD has a positive effect on the methylation (or hypermethylation) on *cg13855288* and *cg19542310* (*LDLRAD3*: Low Density Lipoprotein Receptor Class A Domain Containing 3) in the HIV+ cohort, and a negative effect (or hypomethylation) in the HIV- cohort. On the other hand, in *cg14405266* (*ZNF771*: Zinc Finger Protein 771), *cg14313569* and *cg18609578* (*SPATS2L*: Spermatogenesis Associated Serine Rich 2 Like), COPD has a hypomethylation effect in the HIV+ cohort and a hypermethylation effect in the HIV- cohort.

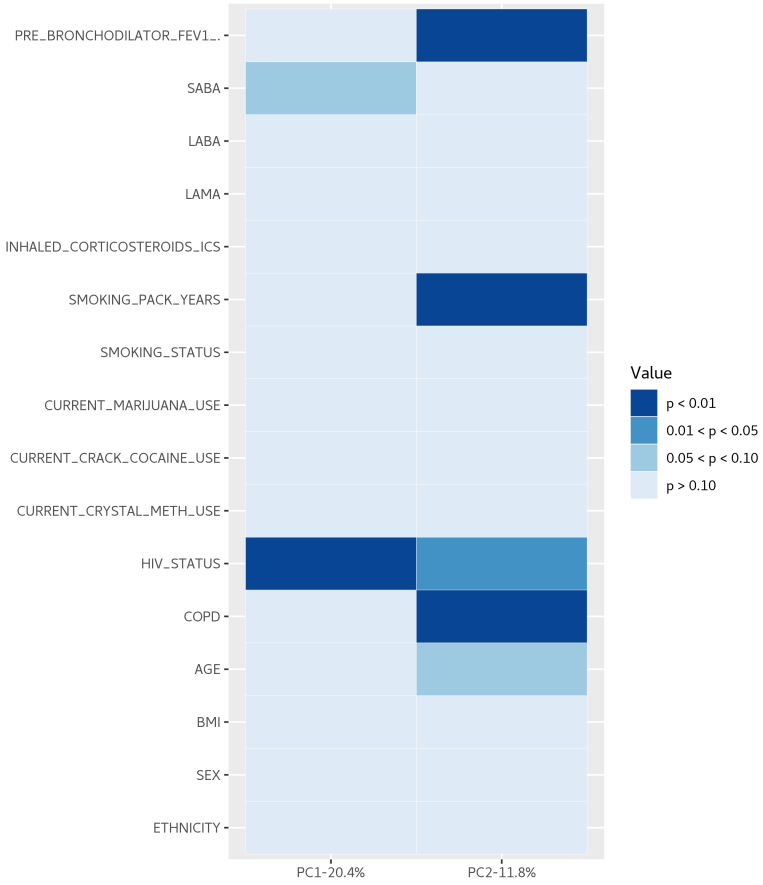
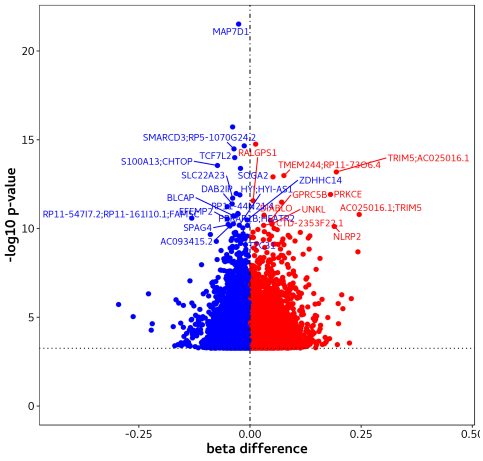
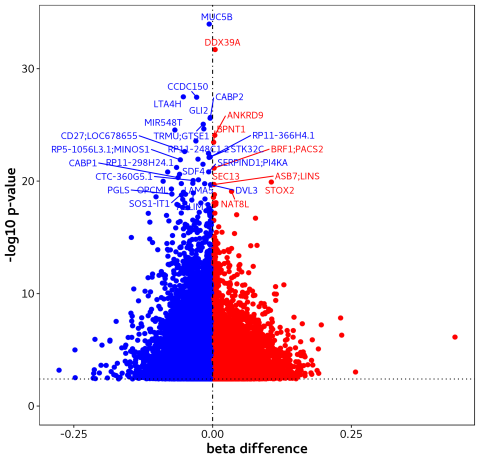


Figure 3.1: Covariate selection for methylome analysis. Heatmap of principal components (PCs) vs. covariates of interest (ethnicity, sex, age, body-mass index (BMI), pre-bronchodilator FEV1%, drug use, smoking status, smoking pack years, inhaled corticosteroid (ICS), long-acting muscarinic antagonists (LAMA), long-acting beta antagonists (LABA) and short-acting beta antagonists (SABA) use, COPD and HIV status). The color scale represents p-values, grey indicating low level and navy indicating high level of significance.



(a)



(b)

Figure 3.2: Volcano plot of the distribution of genes associated with top differentially methylated positions (DMPs) due to (a) the COPD effect, between COPD- and COPD+ (reference) groups, (b) the HIV effect, between HIV- and HIV+ (reference) groups. The x-axes in (a) and (b) represent the effect size difference of the DMPs between subjects with and without COPD and HIV, respectively. The blue and red colours represent hypomethylation and hypermethylation, respectively. The dashed horizontal line represents the -log10 p value that corresponds to the false discovery rate (FDR) < 0.10. The vertical dashed line indicates an absolute beta difference of 0.

Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses identified 41, 106 and 2 pathways, and Gene Ontology (GO) identified 208, 729 and 0 pathways, re-

spectively, that corresponded to the COPD, HIV and COPD*HIV effects (Fig. 3.3). In the COPD-effect analysis, the DMPs were primarily enriched for pathways associated with neurophysiological processes and cell signalling; in the HIV-effect analysis, the most common GO biological pathways were protein metabolism- and cell cycle regulation-related, and KEGG pathways were related to cell signalling, bacterial invasion of cells, and neurophysiological processes; in the COPD*HIV analysis, only two significant KEGG pathways were identified - 'Lysine degradation' and 'Fatty acid biosynthesis'.

Top differentially methylated CpGs at FDR < 0.1									
	Interaction Effect			HIV+		HIV-			
Probe	BetaDiff	P	FDR	COPD BetaDiff	P	COPD BetaDiff	P	Relation to Island	Gene Symbol
cg13855288	0.043	3.27E-10	6.41E-05	0.022	3.45E-06	-0.016	1.404E-04	OpenSea	
cg14405266	-0.019	9.30E-10	1.104E-04	-0.012	1.96E-08	0.008	3.71E-04	SouthShore	ZNF771
cg14313569	-0.058	1.13E-09	1.104E-04	-0.028	3.20E-07	0.033	1.39E-07	OpenSea	
cg19542310	0.025	2.20E-09	1.845E-04	0.014	2.39E-05	-0.013	2.40E-06	OpenSea	LDLRAD3
cg18609578	-0.011	2.35E-09	1.845E-04	-0.006	3.41E-06	0.007	1.42E-05	NorthShore	SPATS2L

Table 3.1: Top 5 CpG sites which have interaction between COPD and HIV and their main effects stratified by HIV. Abbreviations: BetaDiff - Beta Difference.

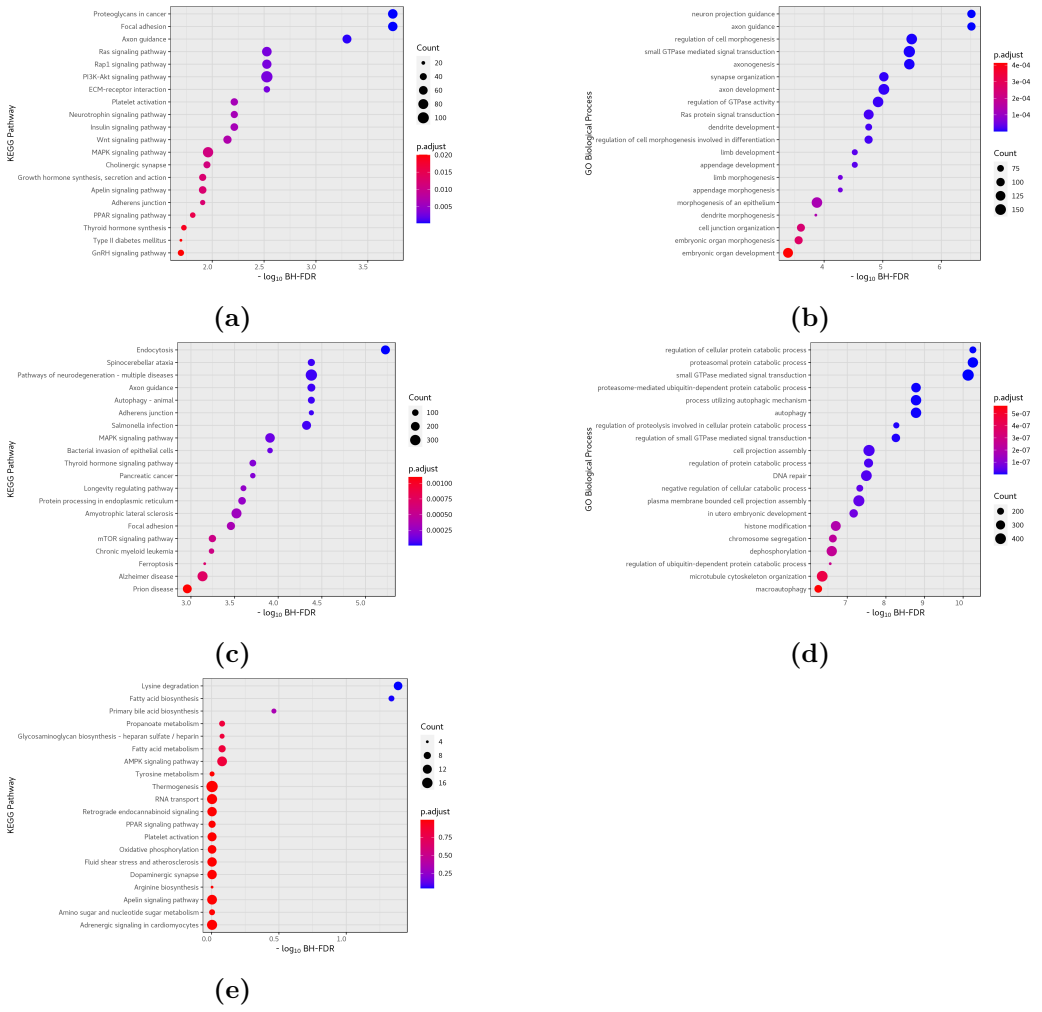


Figure 3.3: Top 20 most significantly enriched KEGG and GO pathways of DMPs associated with the COPD (a, b), HIV (c, d) and COPD*HIV (e) effects; KEGG - column 1 and GO - column 2. (No GO pathways were identified for the COPD*HIV analysis). The size of the circles inside the figure represents the number of overlapping genes characterised by DMPs in the pathways. The colour scale represents level of significance based on p-value. FDR was set to <0.05. Abbreviations: KEGG - Kyoto Encyclopedia of Genes and Genomes; GO - Gene Ontology.

3.5.2 Transcriptome Analyses

PCA used to visualize gene expression patterns among the COPD, HIV and COPD*HIV groups is included in the Appendix (Fig. C.1). Covariate selection analysis identified one variable that may have a significant effect on gene expression - *Sex* (Fig. 3.4). Robust linear regression was applied (correcting for confounder "Sex") and 6,033, 274 and 28 differentially expressed genes (DEGs) with false discovery rate (FDR) < 0.1 were identified in the COPD, HIV, and COPD*HIV analysis, respectively.

As seen in Fig. 3.5, volcano plots showed that genes *AC074143.1*, *SHROOM1* (Shroom Family Member 1) and *PUS3* (Pseudouridine Synthase 3) between the COPD- and COPD+ (reference) groups, and genes *AK1* (Adenylate Kinase 1), *PITX1* (Paired Like Homeodomain 1) and *SAMD15* (Sterile Alpha Motif Domain Containing 15) between the HIV- and HIV+ (reference) groups were identified as the top three DEGs. The top genes identified as significant in the COPD*HIV analyses are in Table. 3.2. COPD has a positive effect on (or upregulates) the expression of genes *CNPY4* (Canopy FGF Signaling Regulator 4) and *AC107021.1* in the HIV+ cohort and a negative effect on (or downregulates)the expression of the same genes in the HIV- cohort. In the opposite direction, COPD has a negative effect on gene expression in the HIV+ cohort and a positive effect on gene expression in the HIV- cohort in the genes *NOP2* (Nucleolar Protein), *KCNE4* (Potassium Voltage-Gated Channel Subfamily E Regulatory Subunit 4) and *RBM17* (RNA Binding Motif Protein 17).

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene enrichment analyses identified 17 and 59 pathways, respectively, associated with the COPD effect. Of these, the top GO Biological Processes included protein localization, energy- and metabolism-related processes, and top KEGG pathways included neurological disease-related pathways, set at FDR < 0.05. No significant pathways were identified in the independent HIV and COPD*HIV analyses (Fig. 3.6).

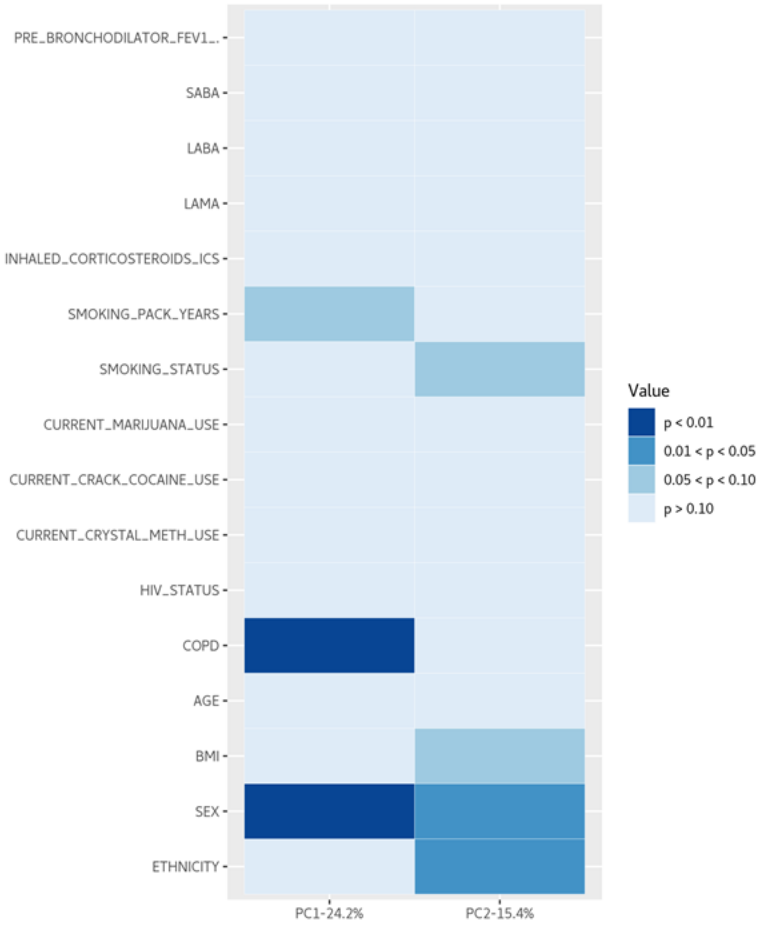
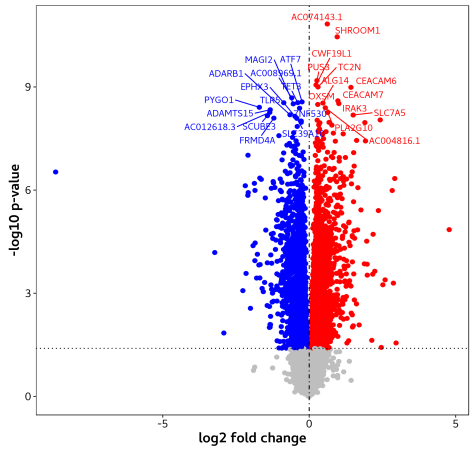
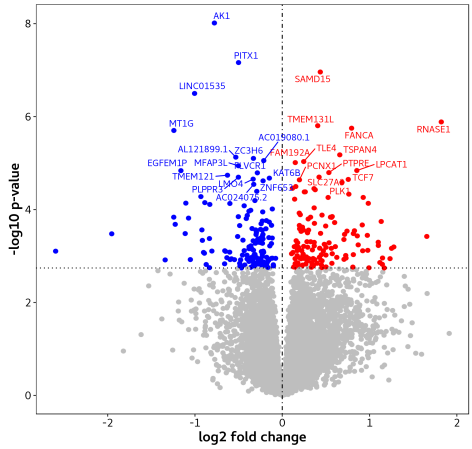


Figure 3.4: Covariate selection for transcriptome analysis. Heatmap of principal components (PCs) vs. covariates of interest (ethnicity, sex, age, body-mass index (BMI), pre-bronchodilator FEV1%, drug use, smoking status, smoking pack years, inhaled corticosteroid (ICS), long-acting muscarinic antagonists (LAMA), long-acting beta antagonists (LABA) and short-acting beta antagonists (SABA) use, COPD and HIV status). The color scale represents p-values, grey indicating low level and navy indicating high level of significance.



(a)

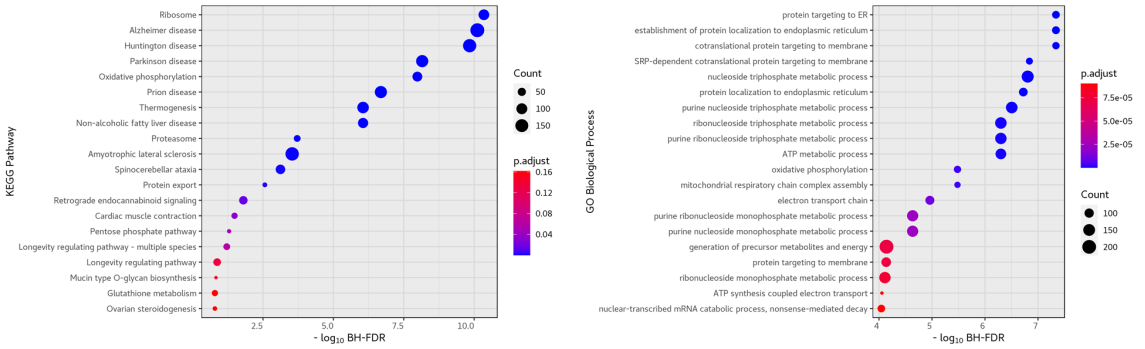


(b)

Figure 3.5: Volcano plot of the distribution of top differentially expressed genes (DEGs) due to (a) the COPD effect, between COPD- and COPD+ (reference) groups, (b) the HIV effect, between HIV- and HIV+ (reference) groups. The x-axes in (a) and (b) represent the effect size difference of the DEGs between subjects with and without COPD and HIV, respectively. The blue and red colours represent downregulated and upregulated genes, respectively. The dashed horizontal line represents the -log10 p value that corresponds to the false discovery rate (FDR)<0.10. The vertical dashed line indicates a log2 fold change of 0.

Top differentially expressed genes at FDR < 0.1								
	Interaction Effect			HIV+		HIV-		
Probe	BetaFC	P	FDR	COPD BetaFC	P	COPD BetaFC	P	Gene
ENSG00000166997.8_4	0.536	1.37E-05	0.074	0.374	4.79E-05	-0.162	0.035	CNPY4
ENSG00000243415.2_6	0.448	6.28E-05	0.082	0.286	2.349E-04	-0.144	0.030	AC107021.1
ENSG00000111641.11_4	-0.304	1.429E-04	0.093	-0.186	3.996E-03	0.124	0.018	NOP2
ENSG00000152049.6_2	-1.184	1.537E-04	0.093	-0.705	3.156E-03	0.513	0.009	KCNE4
ENSG00000134453.16_4	-0.176	1.640E-04	0.095	-0.129	2.100E-03	0.043	0.039	RBM17

Table 3.2: Top 5 genes which have interaction between COPD and HIV and their main effects stratified by HIV (adjusted for sex). Abbreviations: FC - Fold Change.



(a) (b)

Figure 3.6: Top 20 most significantly enriched KEGG and GO pathways of DEGs associated with the COPD effect. (No GO pathways were identified for the HIV effect and COPD*HIV effect analysis). The size of the circles inside the figure represents the number of overlapping DEGs in the pathways. The colour scale represents level of significance based on p-value. FDR was set to <0.05. Abbreviations: KEGG - Kyoto Encyclopedia of Genes and Genomes, GO - Gene Ontology.

3.6 Discussion

From review of literature, we know that alterations in the composition of the HIV airway microbiome may be accompanied by robust changes in host responses such as gene expression as well as regulatory mechanisms such as DNA methylation.

Methylome profiling: We identified top differentially methylated CpG-associated genes like *MAP7D1* (associated with signal transduction) [171] and *SMARCD3* (regulates transcription by altering chromatin structure) [176] between the COPD- and COPD+ groups , and CpG-associated genes *MUC5B* (protects respiratory mucosa from infection and chemical damage by binding to inhaled microorganisms and particles; inhibits HIV-1 entry into target cells by aggregating or trapping the viral particles) [155] [131] and *DDX39A* (involved in RNA metabolic processes such as transcription and translation initiation, splicing and degradation; role in HIV infection cycle by exporting unspliced HIV RNAs from the nucleus) [38] between the HIV- and HIV+ groups.

Our COPD*HIV methylome analysis identified CpG-associated genes such as *LDLRAD3*, which was previously shown to be correlated with bacterial phyla *Proteobacteria* and *Firmicutes* [198]. In previous studies, this gene was found to be down-regulated in COPD [82], and the expressed membrane protein was also important in the characterization of pulmonary-venous endothelial cells localized to the lung parenchyma [178]. In this study, we found that COPD may have a hypermethylating effect on the CpG site of gene *LDLRAD3* in PLWH, thereby potentially downregulating its expression. On the other hand, we also identified the CpG-associated gene *ZNF771*, which regulates transcription by RNA polymerase II, and this gene was also identified to be important in asthma-COPD [123]. In our analysis, we found that COPD had an opposite effect on the methylation of this CpG in PLWH when compared to the HIV-uninfected population.

The top CpG-associated genes in all three analysis were enriched to various KEGG and GO pathways. Most interestingly, the HIV-effect analysis identified DMPs related to pathways related to metabolism and cell cycle regulation and signalling, and the COPD*HIV analysis identified DMPs related to two KEGG pathways - 'Lysine degradation' and 'Fatty acid biosynthesis'. In a recent study by Bowerman *et al* analyzing the fecal microbiome and metabolome of COPD and healthy subjects, amino acid- and lipid-related metabolites were found to be associated with a COPD signature. Although, lysine degradation products, specifically levels of N-acetylcadaverine and its precursor cadaverine, were not significantly different between COPD and healthy samples, they highlighted that a *Rothia* species may be important in N-acetylcadaverine production [21].

On the other hand, impaired fatty acid metabolism has also been implicated in the HIV lung. Cribbs *et al* using BAL samples demonstrated that HIV infection and CD4 count

were related to certain lipid and fatty acid metabolic pathways, which were also linked to microbial families such as *Caulobacteraceae*, *Staphylococcaceae*, *Nocardioidaceae*, and *Streptococcus* [33]. Segal *et al* identified that pulmonary short-chain fatty acids (SCFAs), products of bacterial fermentation, correlated with increased oral anaerobes such as *Prevotella* in the lungs of ART-treated HIV-infected individuals. These SCFAs were also implicated in activating viral replication and inducing persistent Th-17 dysfunction in the host [180]. These results accentuate the connection between microbial colonization and epigenetic changes in the host, and the resultant metabolic and immune outcomes.

Transcriptome profiling: In the COPD-effect analysis, we identified *SHROOM1* (may be linked to actin filament binding; upregulated in COPD+ individuals) [20], and in the HIV-effect analysis, we identified *AK1* (related to HIV-1 replication) [89] to be the top DEGs between the respective groups. Our transcriptomics data looking at the COPD*HIV effect identified the gene *CNPY4*, which was upregulated in PLWH and downregulated in HIV-uninfected individuals as a consequence of the COPD effect. Takahashi *et al* found this gene to be differentially expressed in brushings between current-smokers and non-smokers with severe asthma [200]. The same group also identified the gene *RBM17*, which was upregulated in the HIV-uninfected population and downregulated in PLWH in our analysis. Other top hits included *KCNE4* and *NOP2* an RNA-binding protein. The latter gene has a role in cell cycle progression and chromatin modification, in addition to repressing HIV-1 replication and promoting viral latency [101].

The top genes in the COPD-effect analysis were enriched for biological pathways mainly comprising of protein transport to the endoplasmic reticulum (ER), nucleoside triphosphate metabolic processes, and respiration- and energy-related processes. As the focal hub of cellular metabolism and energetics, the mitochondria consolidates cellular responses and signalling pathways between different organelles, and governs oxidative stress and several age-related processes [116] [36]. At the same time, the ER is crucial for protein processing and transport, regulating cellular redox state, maintaining calcium balance, and the synthesis of lipids, cholesterol, and steroids [53]. During stressful conditions, signals emerging from the ER reach the mitochondria, and subsequently lead to the activation of cell death pathways. In COPD, this mitochondria-ER crosstalk is disrupted by virtue of inhaled toxins [125], indicating its importance in maintaining cellular homeostasis. Our results, which are in line with previous studies, further validate the role of these mitochondria-ER communications.

In summary, the single -omic analyses revealed key ASVs, CpG sites and genes belonging to the microbiome, methylome and the transcriptome, respectively, which may have independent and/or interconnected roles. These results also potentially link the host epigenome, genome to the proteome and metabolome, and set the premise for integrative studies.

Chapter 4

Multiomic Integration

4.1 Introduction

4.1.1 Integrative -omics Approach to Disease

With extensive volumes of microbiome data being generated in recent years, efforts are being taken to recognize how microbial systems interact with each other, their hosts, and features of their environment [3] [16] [88]. However, there are only a limited number of studies that take a 'systems biology' or 'omics' approach in examining how perturbations in the microbiome correlate to distinct aspects of the host biology (e.g., transcriptome, proteome, metabolome, interactome or phenome) [97] [130] [77] [202] [124] [164] [209].

Such studies, simultaneously characterizing the lung microbiome and other -omic profiles in COPD, and in other respiratory diseases in general, are lacking. Most recently, Ramsheh *et al* examined the relationship between the microbiome and transcriptome of mild-to-moderate COPD patients either receiving or not receiving inhaled corticosteroids (ICS) and healthy controls using bronchial brushings. They found that decreased *Prevotella* and increased *Moraxella* levels in COPD were associated with downregulating epithelial defence genes and upregulating pro-inflammatory genes associated with ICS use [160]. Wang *et al* examined the sputum microbiome, transcriptome and proteome in patients with stable COPD and during exacerbations. They found that *Moraxella* and *Haemophilus* had significant associations with interferon and pro-inflammatory signaling pathways in COPD, and that the former was especially important during exacerbations [211]. Sze *et al* analyzed the lung microbiome and host transcriptome in COPD patients and found that *Firmicutes* and *Proteobacteria* were associated with gene expression changes in the host [196]. Cribbs *et al* evaluated the respiratory microbiome-metabolome interaction in PLWH and HIV-uninfected individuals using BAL samples. They identified that specific inflammatory and oxidant metabolic pathways were altered in PLWH with respect to specific lung microbial communities [33]. Such studies further emphasize the need for the comprehensive understanding of the collective host response to the lung microbiome profile, and set the premise for multiomics studies. In this chapter, we performed an integrative analysis by consolidating information from the microbiome with other host -omes such as the methylome and transcriptome.

4.1.2 What is Multiomics?

Multiomics, also called integrated omics, pan-omics, and trans-omics, is the integration of different individual 'omic layers (e.g., genomics, epigenomics, transcriptomics, proteomics), in order to discover novel relationships between the different biological data types, gain insight into mechanisms underlying cellular processes and molecular phenotypes, and obtain a more holistic view of a biological system [54] [103] [94] . Conventionally, individual 'omic layers are analysed independently through univariate statistical methods. However, they fail to capture the crucial relationships between different 'omic features which may influence biological mechanisms and signalling pathways. This can be overcome to a large extent by using multivariate techniques, which can supplement the information obtained from univariate analyses.

Integration of multiomics data has certain limitations. Multiomics studies often carry forward challenges from the individual 'omics datasets, which could undermine further analyses. Individual 'omics data are most commonly generated via different technological platforms, and there is no "gold-standard" workflow addressing issues of data filtering, transformation, normalization and scaling in the individual datasets [133]. This "heterogeneous data" bottleneck may account for major sources of variation in multiomics studies. Another caveat with multiomics is that due to the large volumes of data generated and the use of several complex analytic and statistical tools the results are seldom reproducible. By virtue of that, these studies are computationally expensive and demand high storage space for downloaded files [103].

4.2 Hypothesis

Airway dysbiosis in PLWH with COPD is associated with a host cell injury response.

4.3 Aim 3

Integrate the microbiome, transcriptome and methylome profiles of subjects with COPD and/or HIV to identify key interactions between bacterial ASVs, gene transcripts, and CpG methylation.

4.4 Methods

The microbiome, transcriptome and methylome were statistically integrated using Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO), a multiomics method that simultaneously identifies key omics features among heterogeneous datasets. DIABLO was implemented in the mixOmics R Bioconductor package [113] [166].

The multiomics study includes 76 samples from three types of omics data (for the COPD-effect, HIV-effect and COPD*HIV effect analyses, respectively): microbiome (126 ASVs), transcriptome (6,031, 274 and 28 genes at FDR < 0.1) and methylome (16,206, 92,302 and 4,404 CpGs at FDR < 0.1). All three datasets were normalized and pre-processed according to appropriate omics platforms. A 3×3 symmetric matrix, with values ranging between 0 (indicating no correlation between omics datasets) to 1 (indicating maximum correlation) was used as the design matrix:

	Microbiome	Transcriptome	Methylome
Microbiome	0	0.1	0.1
Transcriptome	0.1	0	0.1
Methylome	0.1	0.1	0

These values were chosen based on a compromise between correlation and discrimination between the features across the different datasets.

A DIABLO model was first fit without variable selection to assess the global performance and choose the number of components (ncomp) for the final DIABLO model. Considering centroids distance measures and the balanced error rates (BER), after 10-fold cross validation repeated 10 times, an optimal number of 2, 3 and 4 components were chosen for the final DIABLO models in the COPD, HIV and COPD*HIV effect analyses, respectively. The datasets were then tuned using sparse partial least squares discriminant analysis (sPLSDA), a component-based integrative approach, to obtain sufficient number of variables for downstream interpretation. The optimal number of variables selected on each component, after 10-fold cross validation repeated 50 times, in the three datasets are given in Table (4.1). Using the chosen parameters, the final DIABLO models were run to identify key interactions between the microbiome, transcriptome, and methylome.

For the COPD*HIV analysis, we also constructed DIABLO circos plots by integrating just the (a) microbiome and methylome, and (b) microbiome and transcriptome, using the same methods as above. These models were run to identify key interactions between just the microbiome and the epigenome, and microbiome and the transcriptome, respectively (Included in Appendix).

	ncomp	Microbiome	Transcriptome	Methylome
COPD-effect	2	(20, 5)	(5, 5)	(25, 10)
HIV-effect	3	(20, 5, 5)	(5, 15, 10)	(25, 5, 5)
COPD*HIV effect	4	(25, 25, 25, 25)	(25, 20, 10, 5)	(15, 10, 20, 20)

Table 4.1: Table showing the optimal number of DIABLO variables selected on each component, after 10-fold cross validation repeated 50 times.

4.5 Results

Integration of the microbiome, transcriptome and methylome was performed to analyze the independent COPD (Fig. 4.1), HIV (Fig. 4.2), and combined COPD*HIV effects (Fig. 4.3).

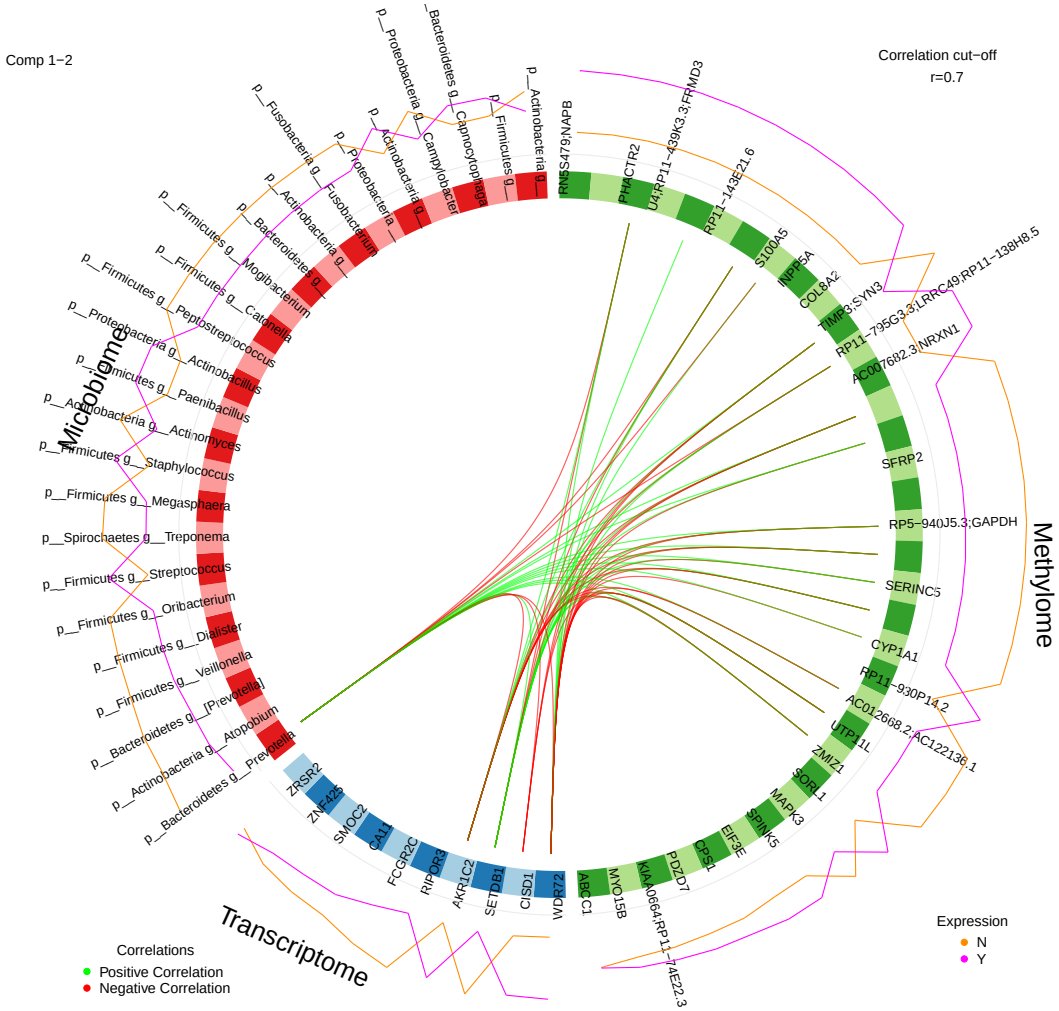


Figure 4.1: DIABLO circos plot showing the within and between correlations between the microbiome, transcriptome and methylome (COPD-effect analysis). The three -omes are represented on the side quadrants; the level of each variable (ASV, CpG and gene) changing with respect to COPD status can be viewed along the circumference. Positive correlation - Green ; Negative correlation - Red; Correlation cutoff = 0.7.

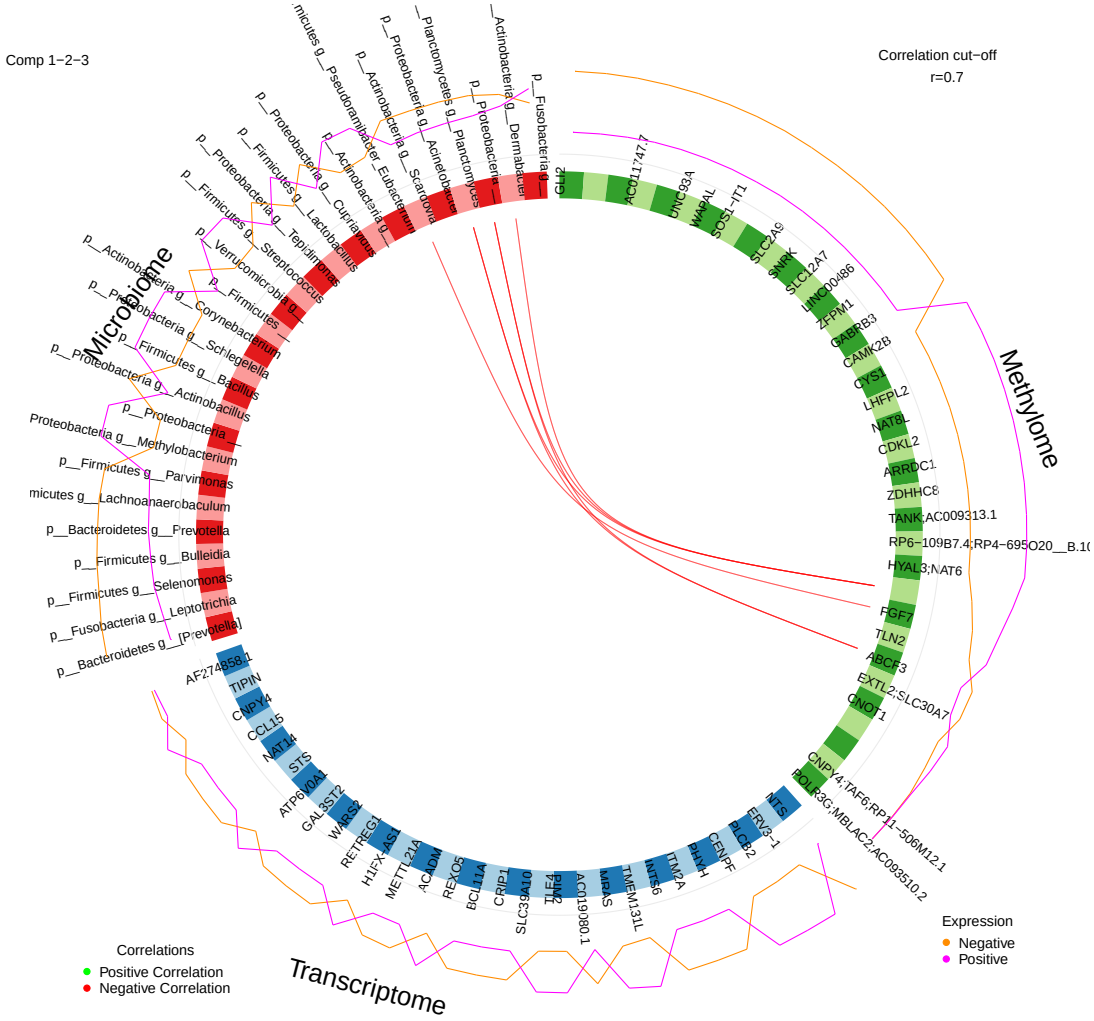


Figure 4.2: DIABLO circos plot showing the within and between correlations between the microbiome, transcriptome and methylome (HIV-effect analysis). The three -omes are represented on the side quadrants; the level of each variable (ASV, CpG and gene) changing with respect to HIV status can be viewed along the circumference. Positive correlation - Green ; Negative correlation - Red; Correlation cutoff = 0.7.

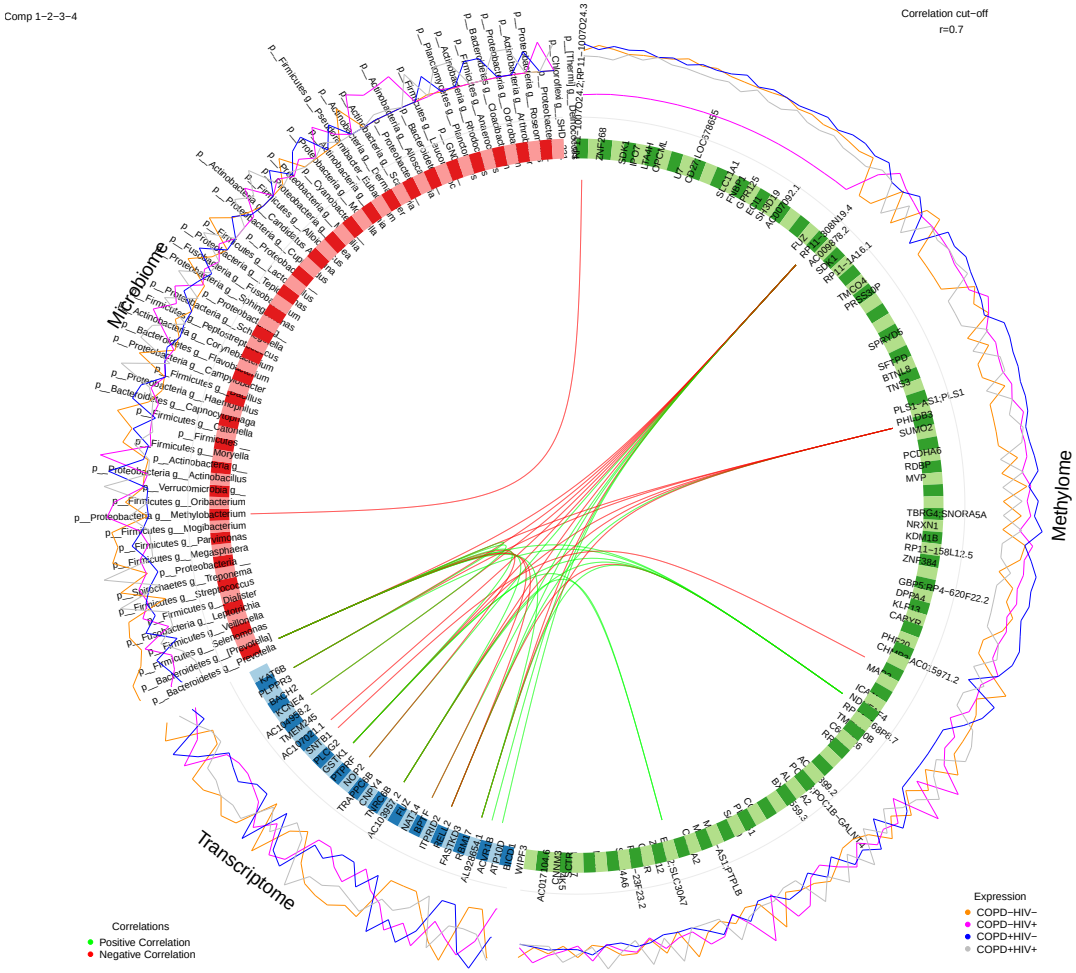


Figure 4.3: DIABLO circos plot showing the within and between correlations between the microbiome, transcriptome and methylome (COPD*HIV-effect analysis). The three -omes are represented on the side quadrants; the level of each variable (ASV, CpG and gene) changing with respect to the combined COPD*HIV status can be viewed along the circumference. Positive correlation - Green ; Negative correlation - Red; Correlation cutoff = 0.7.

Top ASV-Gene Pairs	Correlation
p_Bacteroidetes g_Prevotella - WDR72	-0.797
p_Bacteroidetes g_Prevotella - AKR1C2	-0.765
p_Bacteroidetes g_Prevotella - SETDB1	0.761
Top ASV-CpG Pairs	**Correlation**
p_Bacteroidetes g_Prevotella - CpG TIMP3;SYN3	0.7993
p_Bacteroidetes g_Prevotella - CpG UTP11L	0.791
p_Bacteroidetes g_Prevotella - CpG PHACTR2	-0.760
Top Gene-CpG Pairs	**Correlation**
WDR72 - CpG TIMP3;SYN3	-0.799
WDR72 - CpG UTP11L	-0.784
AKR1C2 - CpG UTP11L	-0.775

Table 4.2: Table showing top ASV-Gene, ASV-CpG and Gene-CpG pairs and their respective correlation values corresponding to the COPD effect. Minus sign indicates negative correlation

The top three ASV-Gene, ASV-CpG and Gene-CpG pairs and their respective correlation values for the COPD, HIV, and COPD*HIV analysis are shown in Tables (4.2), (4.3) and (4.4), respectively. In the HIV analysis, no significant ASV-Gene and Gene-CpG correlations were found at correlation cutoff=0.70. The balanced error rate, used as an estimate of model performance, was ∼30-40% in each component. (Not shown - a secondary DIABLO analysis using a null design (where we assume no correlation between datasets) was performed to see if the model performance could be improved, however no significant improvement was noted).

Top ASV-CpG Pairs	Correlation
p_Actinobacteria g_Scardovia - CpG FGF7	-0.811
p_Proteobacteria - CpG ABCF3	-0.706
p_Planctomycetes g_Planctomyces - CpG ABCF3	-0.706

Table 4.3: Table showing top ASV-CpG pairs and their respective correlation values corresponding to the HIV effect. No ASV-Gene and Gene-CpG correlations were found at correlation cutoff=0.07. Minus sign indicates negative correlation.

Top ASV-Gene Pairs	Correlation
p_Bacteroidetes g_Prevotella - FUZ	0.774
p_Bacteroidetes g_Prevotella - FASTKD3	-0.747
p_Bacteroidetes g_Prevotella - ACVR1B	0.747
Top ASV-CpG Pairs	**Correlation**
p_Bacteroidetes g_Prevotella - CpG FUZ	-0.785
p_Bacteroidetes g_Prevotella - CpG RP11-168P8.7	0.715
p_Bacteroidetes g_Prevotella - CpG PHLDB3	-0.707
Top Gene-CpG Pairs	**Correlation**
FUZ - CpG FUZ	-0.819
ACVR1B - CpG FUZ	-0.800
PTPRF - CpG FUZ	-0.793

Table 4.4: Table showing top ASV-Gene, ASV-CpG and Gene-CpG pairs and their respective correlation values corresponding to the COPD*HIV effect. Minus sign indicates negative correlation.

4.6 Discussion

In disease states, microbial dysbiosis occurs and individual members of the microbiome may be markedly related to host processes such as gene expression and regulation. In our COPD*HIV integration analysis, we identified features belonging to the microbiome, transcriptome and the methylome that may conceivably come together and manifest as disease outcomes.

Our COPD-effect and COPD*HIV integrative 'omics analyses consistently identified the single microbiome feature *Bacteroidetes Prevotella*, whose precise role in lung disease is still widely debated. A host of studies have related *Prevotella* abundance to pulmonary inflammation, especially to the increased expression of Th-17 cytokines [179] [181] [23]. In addition to increased immune stimulation, *Prevotellaceae*-dominated microbial communities have also been found to be enriched for products of branched-chain amino acid metabolism [184].

In complete contrast to these studies, *Prevotella* has also been described in terms of healthy microbial ecosystems. It has been correlated with lung function, exercise capacity, and expression of epithelial genes involved in tight junction promotion and reduced inflammation [160]. This lower immunostimulatory activity of *Prevotella* may be attributed to its cell membrane lipopolysaccharide (LPS) structure [107]. Besides that, *Prevotella* may also interact with other microbes in the respiratory system to exert its effects. One *in vitro* study showed that *Prevotella*, may exhibit its anti-inflammatory properties, by inhibiting the ability of other gram negative bacteria like *Haemophillus influenza* to induce cytokine production [109]. In another study, it was also found to be associated with *Porphyromonas*, with which it can co-aggregate and form heterotrophic biofilms [55]. Many of these properties of *Prevotella* are strain-specific and further investigation into its adaptability and pathogenicity are needed to determine its exact role in health and disease [170].

In our COPD-effect analysis, we found that *Prevotella* was higher in abundance in the COPD- group when compared to the COPD+ group; In the COPD*HIV analysis, we found this ASV to have highest relative abundance in the COPD-HIV- group and lowest in the COPD+HIV+ group. (Not shown - This ASV was also a top hit when we used the DIABLO null design model, where we assume no correlation between datasets (this is different from the design matrix described in our analysis where we assume a 0.1 correlation between datasets)). However, *Prevotella* was not identified to be a top hit in the HIV-effect analysis. These observations reinforce the idea that healthier individuals have higher *Prevotella* and this may be replaced by other microbes with COPD onset and progression.

In our COPD-effect analysis, we also found that *Bacteroidetes Prevotella* was highly correlated with genes *WDR72*, *AKR1C2* and *SETDB1*, and methylation sites *CpG-TIMP3;SYN3*,

CpG-UTP11L and *CpG-PHACTR2*; *CpG-UTP11L* was in turn correlated with gene *WDR72* and *AKR1C2*, and *CpG-TIMP3;SYN3* was correlated with gene *WDR72*. Many of these features have been linked to host processes, for example, gene *AKR1C2* was recently identified as one of the top members in a gene module associated with metabolic processes, responses to oxidative stress and homeostasis [218]. *CpG-TIMP3;SYN3*, associated with gene Tissue Inhibitor of Metalloproteinase 3 (TIMP3) has been previously linked to extracellular matrix (ECM)-binding and metalloproteinase-inhibitory properties [57]. In a molecular context, COPD pathogenesis is reminiscent of degradation of the ECM and upregulation of cellular oxidative stress, both of which are accomplished by various matrix metalloproteinases. These features may thus have an important role to play in regulating oxidative balance and airway remodelling. Other key features include the gene *PHACTR2* (Phosphatase and Actin Regulator 2) related to actin binding and protein phosphatase inhibitor activity, *CpG-UTP11L* related to apoptosis-related gene *UTP11L* (Probable U3 Small Nucleolar RNA-associated Protein 11) [213], and methyltransferase *SETDB1*, increased expression of which was shown to contribute to lung tumorigenesis [165].

Our HIV-effect analysis however identified only significantly correlated ASV-CpG pairs. The top microbiome features included phyla *Actinobacteria*, *Proteobacteria* and *Planctomycetes*, and methylation sites *CpG-ABCF3* (ATP Binding Cassette Subfamily F Member 3) with unknown function, and *CpG-FGF7* (Fibroblast Growth Factor-7), which has a vital role in protecting airway epithelium from oxidant injury that is related to COPD pathogenesis [18].

In our COPD*HIV effect analysis, top microbiome feature *Prevotella* was highly correlated with features of the transcriptome (genes *FASTKD3*, *FUZ* and *ACVR1B*) and the methylome (*CpG-FUZ* and *CpG-PHLDB3*), of which gene *FUZ* and *CpG-FUZ* were also correlated. The gene *FUZ* and its associated CpG site have been implicated in cilium organization and assembly. The expression of this gene has also been associated with the presence of *Firmicutes* and *Proteobacteria* [198]. We found that *FUZ* was hypomethylated and had highest gene expression in the "healthy" COPD-HIV- group, concurrent with its established role in the human body.

Other features include the gene *ACVR1B* that has been noted to have possible relevance to COPD pathogenesis and exacerbations, and may also be related to cGMP-PKG signaling, bacterial invasion of epithelial cells, and actin and immune system-related pathways [26] [144]. Of the 4 groups in the COPD*HIV analysis, this gene showed maximum expression in the COPD-HIV- group, and lowest expression in the COPD+HIV- groups, indicating that COPD may have a role in further driving expression of this gene.

The gene *FASTKD3* and its associated nuclear protein are known to facilitate normal mitochondrial respiration, and also function as a link between RNA translation and the

respiratory machinery [187]. The gene was found to be correlated to *Prevotella* abundance, and its expression was lowest in the COPD-HIV- group and highest in the COPD-HIV+ group, in line with its known function of regulating the energy balance of mitochondria under stress.

CpG-associated gene *PHLDB3* is a target of the tumor suppressor gene p53, which is in turn inactivated in a negative feedback fashion. Under stressful conditions, p53 regulates processes of cell cycle arrest, apoptosis, genomic stability, metabolism and aging [27]. It may be that in HIV-associated COPD, stress factors such as oxidative stress and inflammatory mediators lead to p53 activation, and by suppressing it *PHLDB3* can allow uncontrolled cell division and accelerated aging commonly seen in these patients. The gene *PTPRF* has a role in regulating the assembly and contraction of actin and actomyosin filaments and formation of tight junctions, however, more studies are needed to see if this indeed translates into a barrier function against HIV entry into target cells [215].

This illustrates that interactions between –omes may be critical in driving airflow obstruction seen in PLWH. The microbial feature *Bacteroidetes Prevotella* may be pivotal in its association with host processes such as methylation and gene expression, via the identified candidate transcriptomic and epigenetic features. These features suggest new disease targets involved in pathways regulating mucocilliary clearance, aging/ apoptosis, cellular respiration/ mitochondrial dysfunction and inflammation. Further mechanistic studies can be designed to identify how these targets can be modulated to best improve the clinical outcomes of HIV-associated COPD.

4.7 Challenges

The human microbiome, aside from the well-characterized bacteria, includes archaea, viruses, fungi, and other eukaryotes. These organisms, especially their interactions with one another are critical in disease. We have restricted this study to include only bacteria, however further work encompassing the entire microbial ecosystem is needed.

Besides this inter-kingdom crosstalk, recent studies have indicated that the lung microbiota may be influenced by inter-organ interactions as well [71] [46]. In addition to the microbes entering the lung from the oropharynx and upper respiratory tract through microaspiration, inhalation and mucosal dispersion, there is evidence of inter-compartment crosstalk between the gut and the lungs [182] [44]. These interactions between the gut and lung microbes may be vital for the production of antimicrobial agents, immune modulation, exchange of nutrients, etc. [154]. In this study, we did not analyze gut samples, however, it may be informative to examine the gut microbiome in tandem with that of the lung,

so as to establish an association between possible gut contamination of the lung and the outcomes of disease.

Thirdly, to our knowledge, only few studies have investigated the uniformity of the lung microbiome, although, given that immune infiltration is varied across different compartments of the lung, it may be true that the lung has different microenvironments at various sites [195] [12]. We have addressed this issue to a certain extent by obtaining brushings preferentially from the upper regions of the lung, nevertheless, this could influence our findings. We also acknowledge that evaluating the contribution made by HIV infection per se is complicated by other potential confounding factors such as underlying conditions, and different immune and infiltrating cells in the airways that may affect genomic endpoints. Radiological measures such as computed tomography (CT), and DLCO which are reflective of patterns of altered lung structure and other phenotypic abnormalities may in part account for these changes. Other factors such as sex, age and ART could also have an impact.

Integration of multiomics data also remains a challenging endeavor. Multiomics studies often carry forward challenges from the individual 'omics datasets, which could invalidate further analyses. In this study, data was generated via varied technologies, and was measured on different 'omics technological platforms (i.e., bacterial 16s rRNA and bulk RNA sequencing and microarray). This may be further complicated by the lack of a "gold-standard" workflow addressing issues of data filtering, transformation, normalization and scaling in the individual datasets [133]. To avoid this "heterogeneous data" bottleneck, we examined the individual datasets to identify any major sources of variation, however, we acknowledge that this may be a caveat.

Multiomics integration using DIABLO brings other challenges. The design matrix which aims to optimise the trade-off between correlation and discrimination to identify biologically and clinically relevant 'omics features is user-defined and a fundamental challenge. Striking a balance between leniency and stringency in the choice of design matrix is crucial in uncovering novel multiomics features that have not previously been identified. Secondly, as outlined by the authors of DIABLO, this method assumes that the observed phenotypic response is due to a linear relationship between the 'omics features selected from the individual datasets. However, this assumption of linearity may not hold true in all cases [189]. Another challenge with multiomics is 'over-fitting', which occurs when models are complex and/or over-interpret patterns in the training set, causing the predictive performance to suffer with new data [118]. The hyperparameter tuning step in DIABLO, in particular, runs the risk of over-fitting. We used cross-validation to minimize this, but it does not entirely overcome the over-fitting problem. Future work using an validation dataset can improve this, and provide a more honest assessment of the performance of the model [157]. Despite these challenges, multiomics is gaining in popularity due to its ability to investigate complex mechanisms across molecular layers.

4.8 Future Directions

It is widely accepted that the lung has microenvironments at various sites [12]. In addition to performing microbial community analyses at these unique microenvironments, determining host-derived factors like immune cells/ mediators using single-cell genomic approaches, can allow conclusions into interactions between two main drivers of injury - microbial dysbiosis and inflammation - on the lung epithelium. Host immune responses and disease outcomes may also be manipulated to a large extent by lower airway metabolites, as illustrated by results of our pathway analyses. Hence, including metabolomics data in future integrative -omics studies could have a two-fold effect - (i) connecting disease-related changes in gene expression and microbial dysbiosis to metabolic outcomes in the host, and (ii) settling the "causality dilemma" of whether these metabolites are released as microbial products or whether HIV directs the host metabolism to select for these microbes.

Based on just the 16s rRNA sequencing of bacterial genomes, it is not possible to fully ascertain strain-level resolution, sequence variants, and phenotypes of the different bacterial genera present in the communities. Therefore, future whole-genome sequencing (WGS) and metagenome sequencing, that can improve on the limited taxonomic and functional resolution of 16s rRNA sequencing, can be used to supplement 16s sequencing data. Metagenomics can be extended to survey the viral and fungal populations as well, giving a more comprehensive view of the resident microbiome. In order to identify interactions between these microorganisms, which may work together to facilitate colonization of the lung, microbial correlation networks can also be constructed.

From our analyses and previous reports, it is possible that the microbiome has a critical role to play in the increased susceptibility of PLWH to COPD. Therefore, longitudinal studies evaluating whether the levels of different aerobic and anaerobic microbes change and prevent complications with rehabilitation or improvement, treatment with antimicrobials, ICS, etc., are warranted. We especially highlight *Bacteroidetes Prevotella* as a top ASV; future cell culture models can further help determine the phenotypic response of the microbe in the host.

Bibliography

[1] Rachel I Adams, Ashley C Bateman, Holly M Bik, and James F Meadow. Microbiota of the indoor environment: a meta-analysis. *Microbiome*, 3(1):1–18, 2015.

[2] Aikaterini Alexaki, Yujie Liu, and Brian Wigdahl. Cellular reservoirs of hiv-1 and their role in viral persistence. *Current HIV research*, 6(5):388–400, 2008.

[3] A Paul Alivisatos, MJ Blaser, Eoin L Brodie, Miyoung Chun, Jeffrey L Dangl, Timothy J Donohue, Pieter C Dorrestein, Jack A Gilbert, Jessica L Green, Janet K Jansson, et al. A unified initiative to harness earth's microbiomes. *Science*, 350(6260):507–508, 2015.

[4] Shan V Andrews, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and M Daniele Fallin. "gap hunting" to characterize clustered probe signals in illumina methylation array data. *Epigenetics & chromatin*, 9(1):1–21, 2016.

[5] Engi F Attia, Kathleen M Akgün, Cherry Wongtrakool, Matthew Bidwell Goetz, Maria C Rodriguez-Barradas, David Rimland, Sheldon T Brown, Guy W Soo Hoo, Joon Kim, Patty J Lee, et al. Increased risk of radiographic emphysema in hiv is associated with elevated soluble cd14 and nadir cd4. *Chest*, 146(6):1543–1553, 2014.

[6] JT Attwood, RL Yung, and BC Richardson. Dna methylation and the regulation of gene transcription. *Cellular and Molecular Life Sciences CMLS*, 59(2):241–257, 2002.

[7] Jiu-Wu Bai, Xiao-xin Chen, Shengsheng Liu, Li Yu, and Jin-Fu Xu. Smoking cessation affects the natural history of copd. *International journal of chronic obstructive pulmonary disease*, 12:3323, 2017.

[8] Jennifer M Baker, Kevin J Hinkle, Roderick A McDonald, Christopher A Brown, Nicole R Falkowski, Gary B Huffnagle, and Robert P Dickson. Whole lung tissue is the preferred sampling method for amplicon-based characterization of murine lung microbiota. *Microbiome*, 9(1):1–14, 2021.

[9] Charles RM Bangham. Ctl quality and the control of human retroviral infections. *European journal of immunology*, 39(7):1700–1712, 2009.

[10] PJ Barnes, PG Burney, EK Silverman, BR Celli, J Vestbo, JA Wedzicha, and EF Wouters. Chronic obstructive pulmonary disease. nature reviews. *Disease primers*, 1:15076, 2015.

[11] Christine M Bassis, John R Erb-Downward, Robert P Dickson, Christine M Freeman, Thomas M Schmidt, Vincent B Young, James M Beck, Jeffrey L Curtis, and Gary B Huffnagle. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio*, 6(2):e00037–15, 2015.

[12] Salvatore Battaglia, Thais Mauad, Annemarie M van Schadewijk, Antonia M Vignola, Klaus F Rabe, Vincenzo Bellia, Peter J Sterk, and Pieter S Hiemstra. Differential distribution of inflammatory cells in large and small airways in smokers. *Journal of clinical pathology*, 60(8):907–911, 2007.

[13] Victoria Beasley, Priya V Joshi, Aran Singanayagam, Philip L Molyneaux, Sebastian L Johnston, and Patrick Mallia. Lung microbiology and exacerbations in copd. *International journal of chronic obstructive pulmonary disease*, 7:555, 2012.

[14] James M Beck, Patrick D Schloss, Arvind Venkataraman, Homer Twigg III, Kathleen A Jablonski, Frederic D Bushman, Thomas B Campbell, Emily S Charlson, Ronald G Collman, Kristina Crothers, et al. Multicenter comparison of lung and oral microbiomes of hiv-infected and hiv-uninfected individuals. *American journal of respiratory and critical care medicine*, 192(11):1335–1344, 2015.

[15] Sudha Bhadriraju, Douglas W Fadrosh, Meera K Shenoy, Din L Lin, Kole V Lynch, Kathryn McCauley, Rashida A Ferrand, Edith D Majonga, Grace McHugh, Laurence Huang, et al. Distinct lung microbiota associate with hiv-associated chronic lung disease in children. *Scientific reports*, 10(1):1–9, 2020.

[16] Martin J Blaser, Zoe G Cardon, Mildred K Cho, Jeffrey L Dangl, Timothy J Donohue, Jessica L Green, Rob Knight, Mary E Maxon, Trent R Northen, Katherine S Pollard, et al. Toward a predictive understanding of earth's microbiomes to address 21st century challenges, 2016.

[17] Robert C Bollinger, Ronald S Brookmeyer, Sanjay M Mehendale, Ramesh S Paranjape, Mary E Shepherd, Deepak A Gadkari, and Thomas C Quinn. Risk factors and clinical presentation of acute primary hiv infection in india. *Jama*, 278(23):2085–2089, 1997.

[18] Marcel Bonay, Anne Boutten, Véronique Leçon-Malas, Joëlle Marchal, Paul Soler, Michel Fournier, Guy Leseche, Monique Dehoux, and Bruno Crestani. Hepatocyte and keratinocyte growth factors and their receptors in human lung emphysema. *BMC pulmonary medicine*, 5(1):1–8, 2005.

[19] Gadi Borkow and Zvi Bentwich. Hiv and helminth co-infection: is deworming necessary? *Parasite immunology*, 28(11):605–612, 2006.

[20] Ilse M Boudewijn, Alen Faiz, Katrina Steiling, Erica van der Wiel, Eef D Telenga, Susan JM Hoonhorst, Nick HT Ten Hacken, Corry-Anke Brandsma, Huib AM Kerstjens, Wim Timens, et al. Nasal gene expression differentiates copd from controls and overlaps bronchial gene expression. *Respiratory research*, 18(1):1–10, 2017.

[21] Kate L Bowerman, Saima Firdous Rehman, Annalicia Vaughan, Nancy Lachner, Kurtis F Budden, Richard Y Kim, David LA Wood, Shaan L Gellatly, Shakti D Shukla, Lisa G Wood, et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nature communications*, 11(1):1–15, 2020.

[22] David J Bradshaw, Philip D Marsh, G Keith Watson, and Clive Allison. Role of fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infection and immunity*, 66(10):4729–4732, 1998.

[23] Itzhak Brook and Edith H Frazier. Immune response to fusobacterium nucleatum and prevotella intermedia in the sputum of patients with acute exacerbation of chronic bronchitis. *Chest*, 124(3):832–833, 2003.

[24] Alejandra Cañas-Arboleda, Catalina Hernández-Flórez, Javier Garzón, Claudia Marcela Parra-Giraldo, Juan Felipe Burbano, and José Enrique Cita-Pardo. Colonization by pneumocystis jirovecii in patients with chronic obstructive pulmonary disease: association with exacerbations and lung function status. *Brazilian Journal of Infectious Diseases*, 23:352–357, 2019.

[25] Paul Andres Cardenas, Philip J Cooper, Michael J Cox, Martha Chico, Carlos Arias, Miriam F Moffatt, and William Osmond Cookson. Upper airways microbiota in antibiotic-naive wheezing and healthy infants from the tropics of rural ecuador. *PloS one*, 7(10):e46803, 2012.

[26] Peter J Castaldi, Michael H Cho, Xiaobo Zhou, Weiliang Qiu, Michael Mcgeachie, Bartolome Celli, Per Bakke, Amund Gulsvik, David A Lomas, James D Crapo, et al. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Human molecular genetics*, 24(4):1200–1210, 2015.

[27] Tengfei Chao, Xiang Zhou, Bo Cao, Peng Liao, Hongbing Liu, Yun Chen, Hee-Won Park, Shelya X Zeng, and Hua Lu. Pleckstrin homology domain-containing protein phldb3 supports cancer growth via a negative feedback loop involving p53. *Nature communications*, 7(1):1–12, 2016.

[28] Emily S Charlson, Kyle Bittinger, Andrew R Haas, Ayannah S Fitzgerald, Ian Frank, Anjana Yadav, Frederic D Bushman, and Ronald G Collman. Topographical continuity of bacterial populations in the healthy human respiratory tract. *American journal of respiratory and critical care medicine*, 184(8):957–963, 2011.

[29] Leonard Chávez, Steven Kauder, and Eric Verdin. In vivo, in vitro, and in silico analysis of methylation of the hiv-1 provirus. *Methods*, 53(1):47–53, 2011.

[30] Claudia A Chiriboga, Sharon Fleishman, Susan Champion, Lisa Gaye-Robinson, and Elaine J Abrams. Incidence and prevalence of hiv encephalopathy in children with hiv infection receiving highly active anti-retroviral therapy (haart). *The Journal of pediatrics*, 146(3):402–407, 2005.

[31] Nancy PY Chung, KM Faisal Khan, Robert J Kaner, Sarah L O'Beirne, and Ronald G Crystal. Hiv induces airway basal progenitor cells to adopt an inflammatory phenotype. *Scientific reports*, 11(1):1–13, 2021.

[32] Ana I Hernandez Cordero, Chen Xi Yang, Maen Obeidat, Julia Yang, Julie MacIsaac, Lisa McEwen, David Lin, Michael Kobor, Richard Novak, Fleur Hudson, et al. Dna methylation is associated with airflow obstruction in patients living with hiv. *Thorax*, 2020.

[33] Sushma K Cribbs, Karan Uppal, Shuzhao Li, Dean P Jones, Laurence Huang, Laura Tipton, Adam Fitch, Ruth M Greenblatt, Lawrence Kingsley, David M Guidot, et al. Correlation of the lung microbiota with metabolic profiles in bronchoalveolar lavage fluid in hiv infection. *Microbiome*, 4(1):1–11, 2016.

[34] Kristina Crothers, Adeel A Butt, Cynthia L Gibert, Maria C Rodriguez-Barradas, Stephen Crystal, Amy C Justice, et al. Increased copd among hiv-positive compared to hiv-negative veterans. *Chest*, 130(5):1326–1333, 2006.

[35] Kristina Crothers, Laurence Huang, Joseph L Goulet, Matthew Bidwell Goetz, Sheldon T Brown, Maria C Rodriguez-Barradas, Krisann K Oursler, David Rimland, Cynthia L Gibert, Adeel A Butt, et al. Hiv infection and risk for incident pulmonary diseases in the combination antiretroviral therapy era. *American journal of respiratory and critical care medicine*, 183(3):388–395, 2011.

[36] Charles S Dela Cruz and Min-Jong Kang. Mitochondrial dysfunction and damage associated molecular patterns (damps) in chronic inflammatory diseases. *Mitochondrion*, 41:37–44, 2018.

[37] Lijia Cui, Alison Morris, Laurence Huang, James M Beck, Homer L Twigg III, Erika Von Mutius, and Elodie Ghedin. The microbiome and the lung. *Annals of the American Thoracic Society*, 11(Supplement 4):S227–S232, 2014.

[38] Dirk Daelemans, Elena Afonina, Jakob Nilsson, Gudrun Werner, Jørgen Kjems, Erik De Clercq, George N Pavlakis, and Anne-Mieke Vandamme. A synthetic hiv-1 rev inhibitor interfering with the crm1-mediated nuclear export. *Proceedings of the National Academy of Sciences*, 99(22):14440–14445, 2002.

[39] Angeline T Dang, Sean Cotton, Sumathi Sankaran-Walters, Chin-Shang Li, Chia-Yuan Michael Lee, Satya Dandekar, Bruce J Paster, and Michael D George. Evidence of an increased pathogenic footprint in the lingual microbiome of untreated hiv infected patients. *BMC microbiology*, 12(1):1–10, 2012.

[40] NM Davis, D Proctor, SP Holmes, DA Relman, and BJ Callahan. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. biorxiv, 221499, 2017.

[41] Philip T Diaz, Thomas L Clanton, and Eric R Pacht. Emphysema-like pulmonary disease associated with human immunodeficiency virus infection. *Annals of internal medicine*, 116(2):124–128, 1992.

[42] Philip T Diaz, Mark A King, Eric R Pacht, Mark D Wewers, James E Gadek, David Neal, Haikady N Nagaraja, Janice Drake, and Thomas L Clanton. The pathophysiology of pulmonary diffusion impairment in human immunodeficiency virus infection. *American journal of respiratory and critical care medicine*, 160(1):272–277, 1999.

[43] Philip T Diaz, Mark D Wewers, Eric Pacht, Janice Drake, Haikady N Nagaraja, and Thomas L Clanton. Respiratory symptoms among hiv-seropositive individuals. *Chest*, 123(6):1977–1982, 2003.

[44] Robert P Dickson, John R Erb-Downward, Nicole R Falkowski, Ellen M Hunter, Shanna L Ashley, and Gary B Huffnagle. The lung microbiota of healthy mice are highly variable, cluster by environment, and reflect variation in baseline lung innate immunity. *American journal of respiratory and critical care medicine*, 198(4):497–508, 2018.

[45] Robert P Dickson, John R Erb-Downward, Christine M Freeman, Lisa McCloskey, Nicole R Falkowski, Gary B Huffnagle, and Jeffrey L Curtis. Bacterial topography of the healthy human lower respiratory tract. *MBio*, 8(1), 2017.

[46] Robert P Dickson, Benjamin H Singer, Michael W Newstead, Nicole R Falkowski, John R Erb-Downward, Theodore J Standiford, and Gary B Huffnagle. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nature microbiology*, 1(10):1–9, 2016.

[47] Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

[48] M Bradley Drummond, Laurence Huang, Philip T Diaz, Gregory D Kirk, Eric C Kleerup, Alison Morris, William Rom, Michael D Weiden, Enxu Zhao, Bruce Thompson, et al. Factors associated with abnormal spirometry among hiv-infected individuals. *AIDS (London, England)*, 29(13):1691, 2015.

[49] M Bradley Drummond and Gregory D Kirk. Hiv-associated obstructive lung diseases: insights and implications for the clinician. *The lancet Respiratory medicine*, 2(7):583–592, 2014.

[50] M Bradley Drummond, Gregory D Kirk, Jacquie Astemborski, Mariah M Marshall, Shruti H Mehta, John F McDyer, Robert H Brown, Robert A Wise, and Christian A Merlo. Association between obstructive lung disease and markers of hiv infection in a high-risk cohort. *Thorax*, 67(4):309–314, 2012.

[51] M Bradley Drummond, Christian A Merlo, Jacquie Astemborski, Mariah M Marshall, Annamarie Kisalu, John F Mcdyer, Shruti H Mehta, Robert H Brown, Robert A Wise, et al. The effect of hiv infection on longitudinal lung function decline among injection drug users: a prospective cohort. *AIDS (London, England)*, 27(8):1303, 2013.

[52] Pan Du, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A Kibbe, Lifang Hou, and Simon M Lin. Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*, 11(1):1–9, 2010.

[53] Estefanie Dufey, Denisse Sepúlveda, Diego Rojas-Rivera, and Claudio Hetz. Cellular mechanisms of endoplasmic reticulum stress signaling in health and disease. 1. an overview. *American Journal of Physiology-Cell Physiology*, 307(7):C582–C594, 2014.

[54] Ali Ebrahim, Elizabeth Brunk, Justin Tan, Edward J O'brien, Donghyuk Kim, Richard Szubin, Joshua A Lerman, Anna Lechner, Anand Sastry, Aarash Bordbar, et al. Multi-omic data integration enables discovery of hidden biological regularities. *Nature communications*, 7(1):1–9, 2016.

[55] Marion Engel, David Endesfelder, Brigitte Schloter-Hai, Susanne Kublik, Michael S Granitsiotis, Piera Boschetto, Mariarita Stendardo, Imre Barta, Balazs Dome, Jean-François Deleuze, et al. Influence of lung ct changes in chronic obstructive pulmonary disease (copd) on the human lung microbiome. *PloS one*, 12(7):e0180859, 2017.

[56] John R Erb-Downward, Deborah L Thompson, Meilan K Han, Christine M Freeman, Lisa McCloskey, Lindsay A Schmidt, Vincent B Young, Galen B Toews, Jeffrey L Curtis, Baskaran Sundaram, et al. Analysis of the lung microbiome in the "healthy" smoker and in copd. *PloS one*, 6(2):e16384, 2011.

[57] Dong Fan and Zamaneh Kassiri. Biology of tissue inhibitor of metalloproteinase 3 (timp3), and its therapeutic implications in cardiovascular pathology. *Frontiers in Physiology*, 11:661, 2020.

[58] Emanuele Fanales-Belasio, Mariangela Raimondo, Barbara Suligoi, and Stefano Buttò. Hiv virology and pathogenetic mechanisms of infection: a brief overview. *Annali dell'Istituto superiore di sanita*, 46:5–14, 2010.

[59] Sydney M Finegold, Scot E Dowd, Viktoria Gontcharova, Chengxu Liu, Kathleen E Henley, Randall D Wolcott, Eunseog Youn, Paula H Summanen, Doreen Granpeesheh, Dennis Dixon, et al. Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe*, 16(4):444–453, 2010.

[60] Centers for Disease Control and Prevention. Cdc/ aids and opportunistic infections. `https://www.cdc.gov/hiv/basics/livingwithhiv/opportunisticinfections.html`, 2020.

[61] Emily S Ford, Camille E Puronen, and Irini Sereti. Immunopathogenesis of asymptomatic chronic hiv infection: the calm before the storm. *Current Opinion in HIV and AIDS*, 4(3):206, 2009.

[62] Daniel N Frank, Allison L St Amand, Robert A Feldman, Edgar C Boedeker, Noam Harpaz, and Norman R Pace. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences*, 104(34):13780–13785, 2007.

[63] Wendy S Garrett. Cancer and the microbiota. *Science*, 348(6230):80–86, 2015.

[64] M Patricia George, Mouhamed Kannass, Laurence Huang, Frank C Sciurba, and Alison Morris. Respiratory symptoms and airway obstruction in hiv-infected subjects in the haart era. *PloS one*, 4(7):e6328, 2009.

[65] Dirk Gevers, Subra Kugathasan, Lee A Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, et al. The treatment-naive microbiome in new-onset crohn's disease. *Cell host & microbe*, 15(3):382–392, 2014.

[66] ER Gibney and CM Nolan. Epigenetics and gene expression. *Heredity*, 105(1):4–13, 2010.

[67] Jack A Gilbert, Martin J Blaser, J Gregory Caporaso, Janet K Jansson, Susan V Lynch, and Rob Knight. Current understanding of the human microbiome. *Nature medicine*, 24(4):392–400, 2018.

[68] Matthew R Gingo, M Patricia George, Cathy J Kessinger, Lorrie Lucht, Barbara Rissler, Renee Weinman, William A Slivka, Deborah K McMahon, Sally E Wenzel, Frank C Sciurba, et al. Pulmonary function abnormalities in hiv-infected patients during the current antiretroviral therapy era. *American journal of respiratory and critical care medicine*, 182(6):790–796, 2010.

[69] GOLD. 2020 gold pocket guide. `https://goldcopd.org/wp-content/uploads/2020/03/GOLD-2020-POCKET-GUIDE-ver1.0_FINAL-WMV.pdf`, 2020.

[70] Monica Goldklang and Robert Stockley. Pathophysiology of emphysema and implications. *Chronic Obstructive Pulmonary Diseases*, 3(1):454, 2016.

[71] Alex Grier, Andrew McDavid, Bokai Wang, Xing Qiu, James Java, Sanjukta Bandyopadhyay, Hongmei Yang, Jeanne Holden-Wiltse, Haeja A Kessler, Ann L Gill, et al. Neonatal gut and respiratory microbiota: coordinated development through time and space. *Microbiome*, 6(1):1–19, 2018.

[72] Ellen R Gritz, Damon J Vidrine, Amy B Lazev, Benjamin C Amick, and Roberto C Arduino. Smoking behavior in a low-income multiethnic hiv/aids population. *Nicotine & Tobacco Research*, 6(1):71–77, 2004.

[73] Andrew M Gross, Philipp A Jaeger, Jason F Kreisberg, Katherine Licon, Kristen L Jepsen, Mahdieh Khosroheidari, Brenda M Morsey, Susan Swindells, Hui Shen, Cherie T Ng, et al. Methylome-wide analysis of chronic hiv infection reveals five-year increase in biological age and epigenetic targeting of hla. *Molecular cell*, 62(2):157–168, 2016.

[74] Jessica R Grubb, Anne C Moorman, Rose K Baker, Henry Masur, et al. The changing spectrum of pulmonary disease in patients with hiv infection on antiretroviral therapy. *Aids*, 20(8):1095–1107, 2006.

[75] Jennifer Harrow, Adam Frankish, Jose M Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L Aken, Daniel Barrell, Amonida Zadissa, Stephen Searle, et al. Gencode: the reference human genome annotation for the encode project. *Genome research*, 22(9):1760–1774, 2012.

[76] Mette D Hazenberg, Sigrid A Otto, Birgit HB van Benthem, Marijke ThL Roos, Roel A Coutinho, Joep MA Lange, Dörte Hamann, Maria Prins, and Frank Miedema. Persistent immune activation in hiv-1 infection is associated with progression to aids. *Aids*, 17(13):1881–1888, 2003.

[77] Anna Heintz-Buschart, Patrick May, Cédric C Laczny, Laura A Lebrun, Camille Bellora, Abhimanyu Krishna, Linda Wampach, Jochen G Schneider, Angela Hogan, Carine De Beaufort, et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nature microbiology*, 2(1):1–13, 2016.

[78] Markus Hilty, Conor Burke, Helder Pedro, Paul Cardenas, Andy Bush, Cara Bossley, Jane Davies, Aaron Ervine, Len Poulter, Lior Pachter, et al. Disordered microbial communities in asthmatic airways. *PloS one*, 5(1):e8578, 2010.

[79] Amyn Hirani, Rodrigo Cavallazzi, Tajender Vasu, Monvasi Pachinburavan, Walter K Kraft, Benjamin Leiby, William Short, Joseph Desimone, Kathleen E Squires, Sandra Weibel, et al. Prevalence of obstructive lung disease in hiv population: a cross sectional study. *Respiratory medicine*, 105(11):1655–1661, 2011.

[80] Elaine Y Hsiao, Sara W McBride, Sophia Hsien, Gil Sharon, Embriette R Hyde, Tyler McCue, Julian A Codelli, Janet Chow, Sarah E Reisman, Joseph F Petrosino, et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, 155(7):1451–1463, 2013.

[81] Laurence Huang, Kristina Crothers, Alison Morris, Gena Groner, Melissa Fox, Joan R Turner, Cynthia Merrifield, Shary Eiser, Patrizia Zucchi, and Charles B Beard. Pneumocystis colonization in hiv-infected patients. *The Journal of eukaryotic microbiology*, 50(6):616–617, 2003.

[82] Xinwei Huang, Yunwei Li, Xiaoran Guo, Zongxin Zhu, Xiangyang Kong, Fubing Yu, and Qiang Wang. Identification of differentially expressed genes and signaling pathways in chronic obstructive pulmonary disease via bioinformatic analysis. *FEBS open bio*, 9(11):1880–1899, 2019.

[83] Yaoxing Huang, William A Paxton, Steven M Wolinsky, Avidan U Neumann, Linqi Zhang, Tian He, Stanley Kang, Daniel Ceradini, Zhanqun Jin, Karina Yazdanbakhsh, et al. The role of a mutant ccr5 allele in hiv–1 transmission and disease progression. *Nature medicine*, 2(11):1240–1243, 1996.

[84] Yvonne J Huang, Eugenia Kim, Michael J Cox, Eoin L Brodie, Ron Brown, Jeanine P Wiener-Kronish, and Susan V Lynch. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. *OMICS A Journal of Integrative Biology*, 14(1):9–59, 2010.

[85] Mark W Hull, Peter Phillips, and Julio SG Montaner. Changing global epidemiology of pulmonary manifestations of hiv/aids. *Chest*, 134(6):1287–1298, 2008.

[86] Shoko Iwai, Matthew Fei, Delphine Huang, Serena Fong, Anuradha Subramanian, Katherine Grieco, Susan V Lynch, and Laurence Huang. Oral and airway microbiota in hiv-infected pneumonia patients. *Journal of clinical microbiology*, 50(9):2995–3002, 2012.

[87] University of Michigan Jeffrey L. Curtis, Professor of Internal Medicine. Lung hiv microbiome project (michigan site) (lhmp). `https://clinicaltrials.gov/ct2/show/NCT02392182`, 2016.

[88] Duo Jiang, Courtney R Armour, Chenxiao Hu, Meng Mei, Chuan Tian, Thomas J Sharpton, and Yuan Jiang. Microbiome multi-omics network analysis: statistical considerations, limitations, and opportunities. *Frontiers in genetics*, 10:995, 2019.

[89] Wei-Min Jiang, Xin-Yun Zhang, Yun-Zhi Zhang, Li Liu, and Hong-Zhou Lu. A high throughput rnai screen reveals determinants of hiv-1 activity in host kinases. *International journal of clinical and experimental pathology*, 7(5):2229, 2014.

[90] Bilian Jin, Yajun Li, and Keith D Robertson. Dna methylation: superior or subordinate in the epigenetic hierarchy? *Genes & cancer*, 2(6):607–617, 2011.

[91] W Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007.

[92] Peter A Jones. Functions of dna methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7):484–492, 2012.

[93] James O Kahn and Bruce D Walker. Acute human immunodeficiency virus type 1 infection. *New England Journal of Medicine*, 339(1):33–39, 1998.

[94] Konrad J Karczewski and Michael P Snyder. Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5):299, 2018.

[95] Tyler R Kemnic and Peter G Gulick. Hiv antiretroviral therapy. *StatPearls [Internet]*, 2020.

[96] Victor Kim and Gerard J Criner. Chronic bronchitis and chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 187(3):228–237, 2013.

[97] Gwendoline Kint, Carolina Fierro, Kathleen Marchal, Jos Vanderleyden, and Sigrid CJ De Keersmaecker. Integration of 'omics' data: does it lead to new insights into host–microbe interactions? *Future microbiology*, 5(2):313–328, 2010.

[98] Sam Kint, Wim Trypsteen, Ward De Spiegelaere, Eva Malatinkova, Sabine Kinloch-de Loes, Tim De Meyer, Wim Van Criekinge, and Linos Vandekerckhove. Underestimated effect of intragenic hiv-1 dna methylation on viral transcription in infected individuals. *Clinical epigenetics*, 12(1):1–11, 2020.

[99] Frank Kirchhoff, Thomas C Greenough, Doreen B Brettler, John L Sullivan, and Ronald C Desrosiers. Absence of intact nef sequences in a long-term survivor with nonprogressive hiv-1 infection. *New England Journal of Medicine*, 332(4):228–232, 1995.

[100] Gregory D Kirk and Christian A Merlo. Hiv infection in the etiology of lung cancer: confounding, causality, and consequences. *Proceedings of the American Thoracic Society*, 8(3):326–332, 2011.

[101] Weili Kong, Ayan Biswas, Dawei Zhou, Guillaume Fiches, Koh Fujinaga, Netty Santoso, and Jian Zhu. Nucleolar protein nop2/nsun1 suppresses hiv-1 transcription and promotes viral latency by competing with tat for tar binding and methylation. *PLoS pathogens*, 16(3):e1008430, 2020.

[102] Aleksandar D Kostic, Eunyoung Chun, Lauren Robertson, Jonathan N Glickman, Carey Ann Gallini, Monia Michaud, Thomas E Clancy, Daniel C Chung, Paul Lochhead, Georgina L Hold, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell host & microbe*, 14(2):207–215, 2013.

[103] Michal Krassowski, Vivek Das, Sangram K Sahu, and Biswapriya B Misra. State of the field in multi-omics research: From computational needs to data mining and sharing. *Frontiers in Genetics*, 11, 2020.

[104] Ken M Kunisaki, Jason V Baker, Gary Collins, David M MacDonald, Elzbieta Bakowska, Kamal Marhoum El Filali, Eriobu Nnakelu, Alberto La Rosa, and John E Connett. Lung function decline in early hiv infection: impact of antiretroviral drug timing and drug regimen. *American journal of respiratory and critical care medicine*, 201(6):739–741, 2020.

[105] Umesh G Lalloo, Sandy Pillay, Rosie Mngqibisa, Sabeer Abdool-Gaffar, and Anish Ambaram. Hiv and copd: a conspiracy of risk factors. *Respirology*, 21(7):1166–1172, 2016.

[106] Rafael Laniado-Laborín. Smoking and chronic obstructive pulmonary disease (copd). parallel epidemics of the 21st century. *International journal of environmental research and public health*, 6(1):209–224, 2009.

[107] Jeppe M Larsen, Hanieh S Musavian, Tariq M Butt, Camilla Ingvorsen, Anna H Thysen, and Susanne Brix. Chronic obstructive pulmonary disease and asthma-associated proteobacteria, but not commensal prevotella spp., promote t oll-like receptor 2-independent lung inflammation and pathology. *Immunology*, 144(2):333–342, 2015.

[108] Jeppe Madura Larsen. The immune response to prevotella bacteria in chronic inflammatory disease. *Immunology*, 151(4):363–374, 2017.

[109] Jeppe Madura Larsen, Daniel Bisgaard Steen-Jensen, Janne Marie Laursen, Jonas Nørskov Søndergaard, Hanieh Sadat Musavian, Tariq Mahmood Butt, and Susanne Brix. Divergent pro-inflammatory profile of human dendritic cells in response

to commensal and pathogenic bacteria associated with the airway microbiota. *PLoS One*, 7(2):e31976, 2012.

[110] Anton JM Larsson, Geoff Stanley, Rahul Sinha, Irving L Weissman, and Rickard Sandberg. Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, 15(5):305–307, 2018.

[111] Pui-Pik Law and Michelle L Holland. Dna methylation at the crossroads of gene and environment interactions. *Essays in biochemistry*, 63(6):717–726, 2019.

[112] MB Lawani and A Morris. The respiratory microbiome of hiv-infected individuals. *Expert review of anti-infective therapy*, 14(8):719–729, 2016.

[113] Kim-Anh Lê Cao, Florian Rohart, Ignacio Gonzalez, Sebastien Dejeanand, Benoit key contributors Gautier, Francois Bartolo, et al. mixomics: Omics data integration project. `https://CRAN.R-project.org/package=mixOmics`, 2016.

[114] Elise Orvedal Leiten, Rune Nielsen, Harald Gotten Wiker, Per Sigvald Bakke, Einar Marius Hjellestad Martinsen, Christine Drengenes, Solveig Tangedal, Gunnar Reksten Husebø, and Tomas Mikal Lind Eagan. The airway microbiota and exacerbations of copd. *ERJ open research*, 6(3), 2020.

[115] Annemarie N Lekkerkerker, Yvette v Kooyk, and Teunis BH Geijtenbeek. Viral piracy: Hiv-1 targets dendritic cells for transmission. *Current HIV research*, 4(2):169–176, 2006.

[116] Chad A Lerner, Isaac K Sundar, and Irfan Rahman. Mitochondrial redox system, dynamics, and dysfunction in lung inflammaging and copd. *The international journal of biochemistry & cell biology*, 81:294–306, 2016.

[117] Janice M Leung, Nick Fishbane, Meaghan Jones, Alexander Morin, Stella Xu, Joseph CY Liu, Julie MacIsaac, MJ Milloy, Kanna Hayashi, Julio Montaner, et al. Longitudinal study of surrogate aging measures during human immunodeficiency virus seroconversion. *Aging (Albany NY)*, 9(3):687, 2017.

[118] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: model selection and overfitting. *Nature methods*, 13(9):703–705, 2016.

[119] Yihong Li, Deepak Saxena, Zhou Chen, Gaoxia Liu, Willam R Abrams, Joan A Phelan, Robert G Norman, Gene S Fisch, Patricia M Corby, Floyd Dewhirst, et al. Hiv infection and microbial diversity in saliva. *Journal of clinical microbiology*, 52(5):1400–1411, 2014.

[120] Susan J Little, Angela R McLean, Celsa A Spina, Douglas D Richman, and Diane V Havlir. Viral dynamics of acute hiv-1 infection. *The Journal of experimental medicine*, 190(6):841–850, 1999.

[121] Jason Lloyd-Price, Galeb Abu-Ali, and Curtis Huttenhower. The healthy human microbiome. *Genome medicine*, 8(1):1–11, 2016.

[122] Catherine Lozupone, Adela Cota-Gomez, Brent E Palmer, Derek J Linderman, Emily S Charlson, Erica Sodergren, Makedonka Mitreva, Sahar Abubucker, John Martin, Guohui Yao, et al. Widespread colonization of the lung by tropheryma whipplei in hiv infection. *American journal of respiratory and critical care medicine*, 187(10):1110–1117, 2013.

[123] Mazaher Maghsoudloo, Sadegh Azimzadeh Jamalkandi, Ali Najafi, and Ali Masoudi-Nejad. Identification of biomarkers in common chronic lung diseases by co-expression networks and drug-target interactions analysis. *Molecular Medicine*, 26(1):1–19, 2020.

[124] Tanja V Maier, Marianna Lucio, Lang Ho Lee, Nathan C VerBerkmoes, Colin J Brislawn, Jörg Bernhardt, Regina Lamendella, Jason E McDermott, Nathalie Bergeron, Silke S Heinzmann, et al. Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *MBio*, 8(5), 2017.

[125] Marko Manevski, Thivanka Muthumalage, Dinesh Devadoss, Isaac K Sundar, Qixin Wang, Kameshwar P Singh, Hoshang J Unwalla, Hitendra S Chand, and Irfan Rahman. Cellular stress responses and dysfunctional mitochondrial–cellular senescence, and therapeutics in chronic respiratory diseases. *Redox Biology*, 33:101443, 2020.

[126] Julian R Marchesi and Jacques Ravel. The vocabulary of microbiome research: a proposal, 2015.

[127] Julia L Marcus, Wendy A Leyden, Stacey E Alexeeff, Alexandra N Anderson, Rulin C Hechter, Haihong Hu, Jennifer O Lam, William J Towner, Qing Yuan, Michael A Horberg, et al. Comparison of overall and comorbidity-free life expectancy between insured adults with and without hiv infection, 2000-2016. *JAMA network open*, 3(6):e207954–e207954, 2020.

[128] Mariah M Marshall, Meredith C McCormack, and Gregory D Kirk. Effect of cigarette smoking on hiv acquisition, progression, and mortality. *AIDS Education and Prevention*, 21(3_supplement):28–39, 2009.

[129] Nicola Martin and Quentin Sattentau. Cell-to-cell hiv-1 spread and its implications for immune evasion. *Current Opinion in HIV and AIDS*, 4(2):143–149, 2009.

[130] Ian H McHardy, Maryam Goudarzi, Maomeng Tong, Paul M Ruegger, Emma Schwager, John R Weger, Thomas G Graeber, Justin L Sonnenburg, Steve Horvath, Curtis Huttenhower, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*, 1(1):1–19, 2013.

[131] Isla K McQuaid, Jeffrey R Dorfman, and Anwar S Mall. The comparative inhibitory potency of salivary mucins against human immunodeficiency virus type 1. *Virology*, 553:1–8, 2021.

[132] Lil J Miedzinski. Early clinical signs and symptoms of hiv infection: delaying progression to aids. *Canadian Family Physician*, 38:1401, 1992.

[133] Biswapriya B Misra, Carl Langefeld, Michael Olivier, and Laura A Cox. Integrated omics: tools, advances and future approaches. *Journal of molecular endocrinology*, 62(1):R21–R45, 2019.

[134] DM Mitchell, J Fleming, AJ Pinching, JRW Harris, FM Moss, D Veale, and RJ Shaw. Pulmonary function in human immunodeficiency virus infection. *Am Rev Respir Dis*, 146:745–751, 1992.

[135] Lisa D Moore, Thuc Le, and Guoping Fan. Dna methylation and its basic function. *Neuropsychopharmacology*, 38(1):23–38, 2013.

[136] Alison Morris, James M Beck, Patrick D Schloss, Thomas B Campbell, Kristina Crothers, Jeffrey L Curtis, Sonia C Flores, Andrew P Fontenot, Elodie Ghedin, Laurence Huang, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American journal of respiratory and critical care medicine*, 187(10):1067–1075, 2013.

[137] Alison Morris, Kristina Crothers, James M Beck, and Laurence Huang. An official ats workshop report: Emerging issues and current controversies in hiv-associated pulmonary diseases. *Proceedings of the American Thoracic Society*, 8(1):17–26, 2011.

[138] Alison Morris, M Patricia George, Kristina Crothers, Laurence Huang, Lorrie Lucht, Cathy Kessinger, and Eric C Kleerup. Hiv and chronic obstructive pulmonary disease: is it worse and why? *Proceedings of the American Thoracic Society*, 8(3):320–325, 2011.

[139] Alison Morris and Karen A Norris. Colonization by pneumocystis jirovecii and its role in disease. *Clinical microbiology reviews*, 25(2):297–317, 2012.

[140] Alison Morris, Joseph N Paulson, Hisham Talukder, Laura Tipton, Heather Kling, Lijia Cui, Adam Fitch, Mihai Pop, Karen A Norris, and Elodie Ghedin. Longitudinal analysis of the lung microbiota of cynomolgous macaques during long-term shiv infection. *Microbiome*, 4(1):1–11, 2016.

[141] Alison Morris, Frank C Sciurba, Irina P Lebedeva, Andrew Githaiga, W Mark Elliott, James C Hogg, Laurence Huang, and Karen A Norris. Association of chronic obstructive pulmonary disease severity and pneumocystis colonization. *American journal of respiratory and critical care medicine*, 170(4):408–413, 2004.

[142] Alison Morris, Frank C Sciurba, and Karen A Norris. Pneumocystis: a novel pathogen in chronic obstructive pulmonary disease? *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 5(1):43–51, 2008.

[143] ALISON M MORRIS, LAURENCE HUANG, PETER BACCHETTI, JOAN TURNER, PHILIP C HOPEWELL, JEANNE M WALLACE, PAUL A KVALE, MARK J ROSEN, JEFFREY GLASSROTH, LEE B REICHMAN, et al. Permanent declines in pulmonary function following pneumonia in human immunodeficiency virus-infected persons. *American journal of respiratory and critical care medicine*, 162(2):612–616, 2000.

[144] Jarrett D Morrow, Michael H Cho, John Platig, Xiaobo Zhou, Dawn L DeMeo, Weiliang Qiu, Bartholome Celli, Nathaniel Marchetti, Gerard J Criner, Raphael Bueno, et al. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Human genomics*, 12(1):1–12, 2018.

[145] Kristin N Nelson, Qin Hui, David Rimland, Ke Xu, Matthew S Freiberg, Amy C Justice, Vincent C Marconi, and Yan V Sun. Identification of hiv infection-related dna methylation sites and advanced epigenetic aging in hiv+, treatment-naive us veterans. *AIDS (London, England)*, 31(4):571, 2017.

[146] Francesco Neri, Stefania Rapelli, Anna Krepelova, Danny Incarnato, Caterina Parlato, Giulia Basile, Mara Maldotti, Francesca Anselmi, and Salvatore Oliviero. Intragenic dna methylation prevents spurious transcription initiation. *Nature*, 543(7643):72–77, 2017.

[147] Tamara Newsome, Bing-Jie Li, Nianxiang Zou, and Shyh-Ching Lo. Presence of bacterial phage-like dna sequences in commercial taq dna polymerase reagents. *Journal of clinical microbiology*, 42(5):2264–2267, 2004.

[148] Raymond Niaura, William G Shadel, Kathleen Morrow, Karen Tashima, Timothy Flanigan, and David B Abrams. Human immunodeficiency virus infection, aids, and smoking cessation: the time is now. *Clinical Infectious Diseases*, 31(3):808–812, 2000.

[149] Karen A Norris, Alison Morris, Sangita Patil, and Eustace Fernandes. Pneumocystis colonization, airway inflammation, and pulmonary function decline in acquired immunodeficiency syndrome. *Immunologic research*, 36(1):175–187, 2006.

[150] Divisions of HIV/AIDS Prevention. Hiv and aids faqs — stanford health care. `https://stanfordhealthcare.org/medical-conditions/sexual-and-reproductive-health/hiv-aids/faqs.html#:~:text=How%20well%20does%20HIV%20survive,milk%2C%20saliva%2C%20and%20tears.`

[151] World Health Organization. 2020 gold pocket guide. https://www.who.int/respiratory/copd/burden/en/#:~:text=In%202002%20COPD%20was%20the,leading%20cause%20of%20death%20worldwide., 2020.

[152] World Health Organization. World health organization. https://www.who.int/teams/global-hiv-programme/treatment/hiv-treatment-for-adults, 2020.

[153] Frank J Palella Jr, Kathleen M Delaney, Anne C Moorman, Mark O Loveless, Jack Fuhrer, Glen A Satten, Diane J Aschman, Scott D Holmberg, and HIV Outpatient Study Investigators. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine*, 338(13):853–860, 1998.

[154] Anton Y Peleg, Deborah A Hogan, and Eleftherios Mylonakis. Medically important bacterial–fungal interactions. *Nature reviews microbiology*, 8(5):340–349, 2010.

[155] Juan Perez-Vilar, Scott H Randell, and Richard C Boucher. C-mannosylation of muc5ac and muc5b cys subdomains. *Glycobiology*, 14(4):325–337, 2004.

[156] Charisse Petersen and June L Round. Defining dysbiosis and its influence on host immunity and disease. *Cellular microbiology*, 16(7):1024–1033, 2014.

[157] Milan Picard, Marie-Pier Scott-Boyer, Antoine Bodein, Olivier Périn, and Arnaud Droit. Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal*, 2021.

[158] Ruth Pidsley, Elena Zotenko, Timothy J Peters, Mitchell G Lawrence, Gail P Risbridger, Peter Molloy, Susan Van Djik, Beverly Muhlhausler, Clare Stirzaker, and Susan J Clark. Critical evaluation of the illumina methylationepic beadchip microarray for whole-genome dna methylation profiling. *Genome biology*, 17(1):1–17, 2016.

[159] FA Post, R Wood, and G Maartens. Cd4 and total lymphocyte counts as predictors of hiv disease progression. *QJM: An International Journal of Medicine*, 89(7):505–508, 1996.

[160] Mohammadali Yavari Ramsheh, Koirobi Haldar, Anna Esteve-Codina, Lillie F Purser, Matthew Richardson, Joachim Müller-Quernheim, Timm Greulich, Adam Nowinski, Imre Barta, Mariarita Stendardo, et al. Lung microbiome composition and bronchial epithelial gene expression in patients with copd versus healthy individuals: a bacterial 16s rrna gene sequencing and host transcriptomic analysis. *The Lancet Microbe*, 2021.

[161] Eric J Richards. Inherited epigenetic variation—revisiting soft inheritance. *Nature Reviews Genetics*, 7(5):395–401, 2006.

[162] Charles Rinaldo, Xiao-Li Huang, ZF Fan, Ming Ding, Lisa Beltz, Alison Logar, Dennis Panicali, Gail Mazzara, James Liebmann, and Martin Cottrill. High levels of anti-human immunodeficiency virus type 1 (hiv-1) memory cytotoxic t-lymphocyte activity and low viral load are associated with lack of disease in hiv-1-infected long-term nonprogressors. *Journal of Virology*, 69(9):5838–5842, 1995.

[163] Matthew E Ritchie, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

[164] Richard R Rodrigues, Natalia Shulzhenko, and Andrey Morgun. Transkingdom networks: a systems biology approach to identify causal members of host–microbiota interactions. In *Microbiome Analysis*, pages 227–242. Springer, 2018.

[165] M Rodriguez-Paredes, A Martinez De Paz, L Simo-Riudalbas, S Sayols, C Moutinho, S Moran, A Villanueva, M Vazquez-Cedeira, PA Lazo, F Carneiro, et al. Gene amplification of the histone methyltransferase setdb1 contributes to human lung tumorigenesis. *Oncogene*, 33(21):2807–2813, 2014.

[166] Florian Rohart, Benoît Gautier, Amrit Singh, and Kim-Anh Lê Cao. mixomics: An r package for 'omics feature selection and multiple data integration. *PLoS computational biology*, 13(11):e1005752, 2017.

[167] Mark J Rosen, Yu Lou, Paul A Kvale, A Vijaya Rao, Matthew C Jordan, Albert Miller, Jeffrey Glassroth, Lee B Reichman, Jeanne M Wallace, and Philip C Hopewell. Pulmonary function tests in hiv-infected patients without aids. pulmonary complications of hiv infection study group. *American journal of respiratory and critical care medicine*, 152(2):738–745, 1995.

[168] Eric S Rosenberg, James M Billingsley, Angela M Caliendo, Steven L Boswell, Paul E Sax, Spyros A Kalams, and Bruce D Walker. Vigorous hiv-1-specific cd4+ t cell responses associated with control of viremia. *Science*, 278(5342):1447–1450, 1997.

[169] Justyna Roszkiewicz and Elzbieta Smolewska. Kaleidoscope of autoimmune diseases in hiv infection. *Rheumatology international*, 36(11):1481–1491, 2016.

[170] Yunfeng Ruan, Lu Shen, Yan Zou, Zhengnan Qi, Jun Yin, Jie Jiang, Liang Guo, Lin He, Zijiang Chen, Zisheng Tang, et al. Comparative genome analysis of prevotella intermedia strain isolated from infected root canal reveals features related to pathogenicity and adaptation. *BMC genomics*, 16(1):1–21, 2015.

[171] Dorothy M Ryan, Thomas L Vincent, Jacqueline Salit, Matthew S Walters, Francisco Agosto-Perez, Renat Shaykhiev, Yael Strulovici-Barel, Robert J Downey, Lauren J Buro-Auriemma, Michelle R Staudt, et al. Smoking dysregulates the human airway basal cell transcriptome at copd risk locus 19q13. 2. *PloS one*, 9(2):e88051, 2014.

[172] Andrews S. Fastqc: a quality control tool for high throughput sequence data. `http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`, 2020.

[173] Megan L Salter, Bryan Lau, Vivian F Go, Shruti H Mehta, and Gregory D Kirk. Hiv infection, immune suppression, and uncontrolled viremia are associated with increased multimorbidity among aging injection drug users. *Clinical infectious diseases*, 53(12):1256–1264, 2011.

[174] Sundeep S Salvi and Peter J Barnes. Chronic obstructive pulmonary disease in non-smokers. *The lancet*, 374(9691):733–743, 2009.

[175] AJ Sandford, TD Weir, and P Dp Pare. Genetic risk factors for chronic obstructive pulmonary disease. *European Respiratory Journal*, 10(6):1380–1391, 1997.

[176] Maor Sauler, Maxime Lamontagne, Eric Finnemore, Jose D Herazo-Maya, John Tedrow, Xuchen Zhang, Julia E Morneau, Frank Sciurba, Wim Timens, Peter D Paré, et al. The dna repair transcriptome in severe copd. *European Respiratory Journal*, 52(4), 2018.

[177] Thomas F Schulz, Chris H Boshoff, and Robin A Weiss. Hiv infection and neoplasia. *The Lancet*, 348(9027):587–591, 1996.

[178] Jonas Christian Schupp, Taylor Sterling Adams, Carlos Cosme, Micha Sam Brickman Raredon, Norihito Omote, Sergio Poli, Kadi-Ann Rose, Edward Manning, Maor Sauler, Giuseppe DeIuliis, et al. Integrated single cell atlas of endothelial cells of the human lung. *bioRxiv*, 2020.

[179] Leopoldo N Segal, Alexander V Alekseyenko, Jose C Clemente, Rohan Kulkarni, Benjamin Wu, Hao Chen, Kenneth I Berger, Roberta M Goldring, William N Rom, Martin J Blaser, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome*, 1(1):1–12, 2013.

[180] Leopoldo N Segal, Jose C Clemente, Yonghua Li, Chunhai Ruan, Jane Cao, Mauricio Danckers, Alison Morris, Sarah Tapyrik, Benjamin G Wu, Philip Diaz, et al. Anaerobic bacterial fermentation products increase tuberculosis risk in antiretroviral-drug-treated hiv patients. *Cell host & microbe*, 21(4):530–537, 2017.

[181] Leopoldo N Segal, Jose C Clemente, Jun-Chieh J Tsay, Sergei B Koralov, Brian C Keller, Benjamin G Wu, Yonghua Li, Nan Shen, Elodie Ghedin, Alison Morris, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a th17 phenotype. *Nature microbiology*, 1(5):1–11, 2016.

[182] Leopoldo N Segal, William N Rom, and Michael D Weiden. Lung microbiome for clinicians. new discoveries about bugs in healthy and diseased lungs. *Annals of the American Thoracic Society*, 11(1):108–116, 2014.

[183] Sanjay Sethi. Bacterial infection and the pathogenesis of copd. *Chest*, 117(5):286S–291S, 2000.

[184] Meera K Shenoy, Shoko Iwai, Din L Lin, William Worodria, Irene Ayakaka, Patrick Byanyima, Sylvia Kaswabuli, Serena Fong, Stephen Stone, Emily Chang, et al. Immune response and mortality risk relate to distinct lung microbiomes in patients with hiv and pneumonia. *American journal of respiratory and critical care medicine*, 195(1):104–114, 2017.

[185] Meera K Shenoy and Susan V Lynch. Role of the lung microbiome in hiv pathogenesis. *Current opinion in HIV and AIDS*, 13(1):45–52, 2018.

[186] Timothy W Shipley, Heather M Kling, Alison Morris, Sangita Patil, Jan Kristoff, Siobhan E Guyach, Jessica E Murphy, Xiuping Shao, Frank C Sciurba, Robert M Rogers, et al. Persistent pneumocystis colonization leads to the development of chronic obstructive pulmonary disease in a nonhuman primate model of aids. *The Journal of infectious diseases*, 202(2):302–312, 2010.

[187] Maria Simarro, Alfredo Gimenez-Cassina, Nancy Kedersha, Jean-Bernard Lazaro, Guillaume O Adelmant, Jarrod A Marto, Kirsten Rhee, Sarah Tisdale, Nika Danial, Charaf Benarafa, et al. Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. *Biochemical and biophysical research communications*, 401(3):440–446, 2010.

[188] Viviana Simon, David D Ho, and Quarraisha Abdool Karim. Hiv/aids epidemiology, pathogenesis, prevention, and treatment. *The Lancet*, 368(9534):489–504, 2006.

[189] Amrit Singh, Casey P Shannon, Benoît Gautier, Florian Rohart, Michaël Vacher, Scott J Tebbutt, and Kim-Anh Lê Cao. Diablo: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*, 35(17):3055–3062, 2019.

[190] Rashmi Sinha, Christian C Abnet, Owen White, Rob Knight, and Curtis Huttenhower. The microbiome quality control project: baseline study design and future directions. *Genome biology*, 16(1):1–6, 2015.

[191] Michael W Smith, Michael Dean, Mary Carrington, Cheryl Winkler, Gavin A Huttley, Deborah A Lomb, James J Goedert, Thomas R O'Brien, Lisa P Jacobson, Richard Kaslow, et al. Contrasting genetic influence of ccr2 and ccr5 variants on hiv-1 infection and disease progression. *Science*, 277(5328):959–965, 1997.

[192] Charlotte Soneson, Michael I Love, and Mark D Robinson. Differential analyses for rna-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 2015.

[193] Joan B Soriano and Roberto Rodríguez-Roisin. Chronic obstructive pulmonary disease overview: epidemiology, risk factors, and clinical presentation. *Proceedings of the American Thoracic Society*, 8(4):363–367, 2011.

[194] Bashar Staitieh and David M Guidot. Noninfectious pulmonary complications of human immunodeficiency virus infection. *The American journal of the medical sciences*, 348(6):502–511, 2014.

[195] Marc A Sze, Pedro A Dimitriu, Shizu Hayashi, W Mark Elliott, John E McDonough, John V Gosselink, Joel Cooper, Don D Sin, William W Mohn, and James C Hogg. The lung tissue microbiome in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 185(10):1073–1080, 2012.

[196] Marc A Sze, Pedro A Dimitriu, Masaru Suzuki, John E McDonough, Josh D Campbell, John F Brothers, John R Erb-Downward, Gary B Huffnagle, Shizu Hayashi, W Mark Elliott, et al. Host response to the lung microbiome in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 192(4):438–445, 2015.

[197] Marc A Sze, Stella Xu, Janice M Leung, Emily A Vucic, Tawimas Shaipanich, Aida Moghadam, Marianne Harris, Silvia Guillemi, Sunita Sinha, Corey Nislow, et al. The bronchial epithelial cell bacterial microbiome and host response in patients infected with human immunodeficiency virus. *BMC pulmonary medicine*, 16(1):1–10, 2016.

[198] Marc Alexander Sze. *The bacterial lung tissue microbiome in the pathogenesis of chronic obstructive pulmonary disease.* PhD thesis, University of British Columbia, 2015.

[199] Brown TA. *Genomes. 2nd edition. Oxford: Wiley-Liss; 2002. Chapter 3, Transcriptomes and Proteomes.* 2002.

[200] Kentaro Takahashi, Stelios Pavlidis, Francois Ng Kee Kwong, Uruj Hoda, Christos Rossios, Kai Sun, Matthew Loza, Fred Baribaud, Pascal Chanez, Steve J Fowler, et al. Sputum proteomics and airway cell transcripts of current and ex-smokers with severe asthma in u-biopred: an exploratory analysis. *European respiratory journal*, 51(5), 2018.

[201] Andrew E Teschendorff, Francesco Marabita, Matthias Lechner, Thomas Bartlett, Jesper Tegner, David Gomez-Cabrero, and Stephan Beck. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 k dna methylation data. *Bioinformatics*, 29(2):189–196, 2013.

[202] Casey M Theriot, Mark J Koenigsknecht, Paul E Carlson, Gabrielle E Hatton, Adam M Nelson, Bo Li, Gary B Huffnagle, Jun Z Li, and Vincent B Young.

Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to clostridium difficile infection. *Nature communications*, 5(1):1–10, 2014.

[203] Yuan Tian, Tiffany J Morris, Amy P Webster, Zhen Yang, Stephan Beck, Andrew Feber, and Andrew E Teschendorff. Champ: updated methylation analysis pipeline for illumina beadchips. *Bioinformatics*, 33(24):3982–3984, 2017.

[204] Timothy J Triche Jr, Daniel J Weisenberger, David Van Den Berg, Peter W Laird, and Kimberly D Siegmund. Low-level processing of illumina infinium dna methylation beadarrays. *Nucleic acids research*, 41(7):e90–e90, 2013.

[205] M Triplette, K Crothers, and EF Attia. Non-infectious pulmonary diseases and hiv. *Current HIV/AIDS Reports*, 13(3):140–148, 2016.

[206] Aurélien Trompette, Eva S Gollwitzer, Koshika Yadava, Anke K Sichelstiel, Norbert Sprenger, Catherine Ngom-Bru, Carine Blanchard, Tobias Junt, Laurent P Nicod, Nicola L Harris, et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature medicine*, 20(2):159–166, 2014.

[207] Homer L Twigg III, Kenneth S Knox, Jin Zhou, Kristina A Crothers, David E Nelson, Evelyn Toh, Richard B Day, Huaiying Lin, Xiang Gao, Qunfeng Dong, et al. Effect of advanced hiv infection on the respiratory microbiome. *American journal of respiratory and critical care medicine*, 194(2):226–235, 2016.

[208] Claire Vandergeeten, Rémi Fromentin, and Nicolas Chomont. The role of cytokines in the establishment, persistence and eradication of the hiv reservoir. *Cytokine & growth factor reviews*, 23(4-5):143–149, 2012.

[209] Qing Wang, Kaicen Wang, Wenrui Wu, Eleni Giannoulatou, Joshua WK Ho, and Lanjuan Li. Host and microbiome multi-omics integration: applications and methodologies. *Biophysical reviews*, 11(1):55–65, 2019.

[210] Zhang Wang, Barbara Maschera, Simon Lea, Umme Kolsum, David Michalovich, Stephanie Van Horn, Christopher Traini, James R Brown, Edith M Hessel, and Dave Singh. Airway host-microbiome interactions in chronic obstructive pulmonary disease. *Respiratory research*, 20(1):1–14, 2019.

[211] Zhang Wang, Barbara Maschera, Simon Lea, Umme Kolsum, David Michalovich, Stephanie Van Horn, Christopher Traini, James R Brown, Edith M Hessel, and Dave Singh. Airway host-microbiome interactions in chronic obstructive pulmonary disease. *Respiratory research*, 20(1):1–14, 2019.

[212] James R Waymack and Vidya Sundareshan. Acquired immune deficiency syndrome. 2020.

[213] Helen L Wright, Huw B Thomas, Robert J Moots, and Steven W Edwards. Rna-seq reveals activation of both common and cytokine-specific pathways following neutrophil priming. *PloS one*, 8(3):e58598, 2013.

[214] Stella Xu, Amy Tsai, Marc A Sze, Emily A Vucic, Tawimas Shaipanich, Marianne Harris, Silvia Guillemi, Julia Yang, Sunita Sinha, Corey Nislow, et al. Decreased microbiome diversity in the hiv small airway epithelium. *Respiratory research*, 19(1):1–8, 2018.

[215] Wen-Wen Xu, Miao-Jun Han, Dai Chen, Ling Chen, Yan Guo, Andrew Willden, Di-Qiu Liu, and Hua-Tang Zhang. Genome-wide search for the genes accountable for the induced resistance to hiv-1 infection in activated cd4+ t cells: apparent transcriptional signatures, co-expression networks and possible cellular processes. *BMC medical genomics*, 6(1):1–16, 2013.

[216] Yan Xu. *An exploration of the lung microbiome and DNA methylation in patients infected with human immunodeficiency virus*. PhD thesis, University of British Columbia, 2016.

[217] Chen Xi Yang, Emma Schon, Ma'en Obeidat, Michael S Kobor, Lisa McEwen, Julie MacIsaac, David Lin, Richard M Novak, Fleur Hudson, Hartwig Klinker, et al. Occurrence of accelerated epigenetic aging and methylation disruptions in human immunodeficiency virus infection before antiretroviral therapy. *The Journal of Infectious Diseases*, 2020.

[218] Chen Xi Yang, Henry Shi, Irving Ding, Stephen Milne, Ana I Hernandez Cordero, Cheng Wei Tony Yang, Edward Kyoo-Hoon Kim, Tillie-Louise Hackett, Janice Leung, Don D Sin, et al. Widespread sexual dimorphism in the transcriptome of human airway epithelium in response to smoking. *Scientific reports*, 9(1):1–14, 2019.

[219] Libing Yang, Daniel G Dunlap, Shulin Qin, Adam Fitch, Kelvin Li, Carl D Koch, Mehdi Nouraie, Rebecca DeSensi, Ken S Ho, Jeremy J Martinson, et al. Alterations in oral microbiota in hiv are related to decreased pulmonary function. *American journal of respiratory and critical care medicine*, 201(4):445–457, 2020.

[220] Shun-Wei Julia Yang. *Microbiota and telomere shortening in gut-lung axis of human immunodeficiency virus infected donors*. PhD thesis, University of British Columbia, 2018.

[221] Peilin Zheng, Zhixia Li, and Zhiguang Zhou. Gut microbiome in type 1 diabetes: A comprehensive review. *Diabetes/metabolism research and reviews*, 34(7):e3043, 2018.

[222] He Zhou, Lin Sun, Siwen Zhang, Xue Zhao, Xiaokun Gang, and Guixia Wang. Evaluating the causal role of gut microbiota in type 1 diabetes and its possible pathogenic mechanisms. *Frontiers in endocrinology*, 11:125, 2020.

Appendix A

Microbiome Analysis

	Group 1	Group 2	†Adjusted p-values	
			Shannon Index	Faith PD
COPD effect	COPD-	COPD+	0.001	0.0004
HIV effect	HIV-	HIV+	0.002	0.003
COPD*HIV effect	COPD+HIV+	COPD+HIV-	0.050	0.017
	COPD+HIV+	COPD-HIV+	0.094	0.017
	COPD+HIV+	COPD-HIV-	0.0001	0.0001
	COPD+HIV-	COPD-HIV+	0.952	0.657
	COPD+HIV-	COPD-HIV-	0.014	0.017
	COPD-HIV+	COPD-HIV-	0.047	0.089

Table A.1: Pairwise Kruskal-Wallis test of alpha diversity measured using the Shannon Index and Faith PD metrics. †Adjusted p-values for multiple comparisons between groups was obtained using the Benjamini-Hochberg procedure (false discovery rate method correction). Definition of abbreviations: PD – Phylogenetic Diversity.

	Group 1	Group 2	Sample size	p-value	†q-value
COPD effect	COPD-	COPD+	76	0.001	0.001
HIV effect	HIV-	HIV+	76	0.01	0.01
COPD*HIV effect	COPD+HIV+	COPD+HIV-	40	0.081	0.0972
	COPD+HIV+	COPD-HIV+	34	0.012	0.024
	COPD+HIV+	COPD-HIV-	38	0.001	0.006
	COPD+HIV-	COPD-HIV+	38	0.119	0.119
	COPD+HIV-	COPD-HIV-	42	0.009	0.024
	COPD-HIV+	COPD-HIV-	36	0.019	0.0285

Table A.2: Pairwise PERMANOVA comparisons based on COPD, HIV and COPD*HIV statuses. †q-values obtained using Benjamini-Hochberg FDR correction.

Phylum	COPD- (n=36)	COPD+ (n=40)	*p-value	†Adjusted p-value
Firmicutes	0.386[0.221]	0.4[0.342]	0.357	0.536
Bacteroidetes	0.418[0.314]	0.185[0.355]	0.003	0.018
Proteobacteria	0.065[0.098]	0.169[0.246]	0.042	0.084
Actinobacteria	0.031[0.054]	0.037[0.067]	0.815	0.874
Fusobacteria	0.018[0.03]	0[0.025]	0.012	0.036
Verrucomicrobia	0[0.002]	0[0.006]	0.874	0.874

Table A.3: Relative taxa abundance comparisons at the phylum level between the COPD+ and COPD- groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Mann–Whitney U-test; †Adjusted p-values obtained using the Benjamini-Hochberg procedure (False Discovery Rate method). Definition of abbreviations: AEC – Airway epithelial cells.

Genus	COPD- (n=36)	COPD+ (n=40)	*p-value	†Adjusted p-value
Prevotella∗	0.334[0.302]	0.144[0.271]	0.0008	0.010
Veillonella	0.154[0.178]	0.065[0.156]	0.011	0.027
Streptococcus	0.057[0.060]	0.094[0.196]	0.233	0.028
Haemophilus	0.015[0.049]	0.001[0.068]	0.157	0.209
Paenibacillus	0.000[0.003]	0.005[0.020]	0.028	0.042
Rothia	0.008[0.025]	0.006[0.026]	0.680	0.742
Actinobacillus	0.000[0.000]	0.000[0.000]	0.980	0.980
Megasphaera	0.012[0.046]	0.000[0.008]	0.005	0.024
Prevotella∓	0.012[0.043]	0.000[0.011]	0.008	0.024
Neisseria	0.002[0.019]	0.000[0.004]	0.158	0.032
Selenomonas	0.002[0.018]	0.000[0.001]	0.020	0.035
Fusobacterium	0.008[0.027]	0.000[0.015]	0.007	0.024

Table A.4: Relative taxa abundance comparisons at the genus level between the COPD+ and COPD- groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Mann–Whitney U-test; †Adjusted p-values obtained using the Benjamini-Hochberg procedure (False Discovery Rate method). *Prevotella*∗ - *Prevotella*[f-*Prevotellaceae*], *Prevotella*∓ - *Prevotella*[f-*Paraprevotellaceae*]. Definition of abbreviations: AEC – Airway epithelial cells.

Phylum	Relative Abundance	
	Median [IQR]	Range
Firmicutes	0.398 [0.277]	0.007 - 0.983
Bacteroidetes	0.29 [0.42]	0 - 0.645
Proteobacteria	0.107 [0.206]	0.001 - 0.942
Actinobacteria	0.033 [0.059]	0 - 0.989
Fusobacteria	0.088 [0.034]	0 - 0.173
Verrucomicrobia	0 [0.004]	0 - 0.385
Others	0 [0]	0 - 0.663

Table A.5: Relative abundance of most abundant phyla (average relative abundance $\geq$ 2%) observed in AEC samples. Definition of abbreviations: AEC – Airway epithelial cells; IQR - Interquartile range.

Genus	Relative Abundance	
	Median [IQR]	Range
Prevotella∗	0.244 [0.358]	0 - 0.576
Veillonella	0.122 [0.179]	0 - 0.389
Streptococcus	0.067 [0.139]	0 - 0.728
Haemophilus	0.009 [0.053]	0 - 0.942
Paenibacillus	0.001 [0.013]	0 - 0.814
Rothia	0.008 [0.027]	0 - 0.232
Actinobacillus	0 [0]	0 - 0.646
Megasphaera	0 [0.031]	0 - 0.199
Prevotella∓	0.001 [0.025]	0 - 0.122
Neisseria	0 [0.008]	0 - 0.306
Selenomonas	0 [0.009]	0 - 0.129
Fusobacterium	0.003 [0.024]	0 - 0.146
Others	0 [0]	0 - 0.989

Table A.6: Relative abundance of most abundant genera (average relative abundance $\geq$ 1%) observed in AEC samples. Definition of abbreviations: AEC – Airway epithelial cells; IQR - Interquartile range. *Prevotella*∗ - *Prevotella*[f-*Prevotellaceae*], *Prevotella*∓ - *Prevotella*[f-*Paraprevotellaceae*].

Phylum	HIV- (n=42)	HIV+ (n=34)	*p-value	†Adjusted p-value
Firmicutes	0.362[0.19]	0.458[0.347]	0.123	0.247
Bacteroidetes	0.364[0.361]	0.203[0.365]	0.034	0.103
Proteobacteria	0.11[0.265]	0.08[0.179]	0.605	0.623
Actinobacteria	0.032[0.052]	0.038[0.071]	0.623	0.623
Fusobacteria	0.022[0.043]	0[0.011]	0.001	0.008
Verrucomicrobia	0[0.002]	0[0.016]	0.276	0.414

Table A.7: Relative taxa abundance comparisons at the phylum level between the HIV+ and HIV- groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Mann–Whitney U-test; †Adjusted p-values obtained using the Benjamini-Hochberg procedure (False Discovery Rate method). Definition of abbreviations: AEC – Airway epithelial cells.

Genus	HIV- (n=42)	HIV+ (n=34)	*p-value	†Adjusted p-value
Prevotella∗	0.295[0.305]	0.127[0.315]	0.020	0.048
Veillonella	0.132[0.114]	0.112[0.268]	0.996	0.996
Streptococcus	0.067[0.081]	0.067[0.164]	0.627	0.752
Haemophilus	0.013[0.054]	0.002[0.051]	0.354	0.472
Paenibacillus	0.000[0.012]	0.001[0.020]	0.817	0.892
Rothia	0.010[0.021]	0.002[0.028]	0.105	0.158
Actinobacillus	0.000[0.000]	0.000[0.000]	0.030	0.061
Megasphaera	0.008[0.036]	0.000[0.007]	0.071	0.122
Prevotella∓	0.012[0.042]	0.000[0.001]	0.0003	0.004
Neisseria	0.002[0.016]	0.000[0.002]	0.017	0.048
Selenomonas	0.001[0.039]	0.000[0.001]	0.014	0.048
Fusobacterium	0.010[0.027]	0.000[0.006]	0.012	0.048

Table A.8: Relative taxa abundance comparisons at the genus level between the HIV+ and HIV- groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Mann–Whitney U-test; †Adjusted p-values obtained using the Benjamini-Hochberg procedure (False Discovery Rate method). *Prevotella*∗ - *Prevotella*[f-*Prevotellaceae*], *Prevotella*∓ - *Prevotella*[f-*Paraprevotellaceae*]. Definition of abbreviations: AEC – Airway epithelial cells.

Phylum	COPD-HIV- (n=20)	COPD-HIV+ (n=16)	COPD+HIV- (n=22)	COPD+HIV+ (n=18)	*p-value	†Effect size
Firmicutes	0.375[0.170]	0.386[0.379]	0.362[0.267]	0.541[0.359]	0.090	0.048[small]
Bacteroidetes	0.472[0.183]	0.273[0.352]	0.279[0.349]	0.107[0.333]	0.004	0.143[large]
Proteobacteria	0.067[0.116]	0.061[0.080]	0.232[0.287]	0.155[0.166]	0.198	0.023[small]
Actinobacteria	0.026[0.026]	0.054[0.081]	0.047[0.097]	0.022[0.048]	0.121	0.039[small]
Fusobacteria	0.030[0.044]	0.007[0.015]	0.009[0.044]	0.000[0.003]	0.0007	0.194[large]
Verrucomicrobia	0.000[0.002]	0.000[0.004]	0.000[0.002]	0.000[0.032]	0.523	-0.011[small]

Table A.9: Relative taxa abundance comparisons at the phylum level between the COPD-HIV-, COPD-HIV+, COPD+HIV- and COPD+HIV+ groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Kruskal-Wallis test; †Effect size calculated based on Kruskal-Wallis H-statistic [Eta squared method]. Definition of abbreviations: AEC – Airway epithelial cells.

Genus	COPD-HIV- (n=20)	COPD-HIV+ (n=16)	COPD+HIV- (n=22)	COPD+HIV+ (n=18)	*p-value	†Effect size
Prevotella∗	0.414[0.151]	0.228[0.308]	0.214[0.286]	0.024[0.240]	0.0008	0.191[large]
Veillonella	0.157[0.106]	0.149[0.221]	0.073[0.119]	0.040[0.256]	0.077	0.0535[small]
Streptococcus	0.054[0.053]	0.061[0.147]	0.080[0.140]	0.140[0.253]	0.588	-0.0150[small]
Haemophilus	0.015[0.046]	0.014[0.048]	0.005[0.069]	0.000[0.055]	0.331	0.00590[small]
Paenibacillus	0.000[0.005]	0.000[0.001]	0.001[0.014]	0.009[0.064]	0.024	0.0901[moderate]
Rothia	0.008[0.011]	0.004[0.040]	0.012[0.050]	0.001[0.023]	0.401	-0.000852[small]
Actinobacillus	0.000[0.000]	0.000[0.000]	0.000[0.000]	0.000[0.000]	0.100	0.0452[small]
Megasphaera	0.032[0.054]	0.005[0.037]	0.000[0.022]	0.000[0.000]	0.011	0.112[moderate]
Prevotella∓	0.026[0.041]	0.000[0.016]	0.004[0.021]	0.000[0.000]	0.0002	0.236[large]
Neisseria	0.003[0.027]	0.000[0.009]	0.000[0.007]	0.000[0.000]	0.009	0.119[moderate]
Selenomonas	0.005[0.046]	0.000[0.004]	0.000[0.024]	0.000[0.000]	0.010	0.116[moderate]
Fusobacterium	0.025[0.032]	0.003[0.010]	0.002[0.020]	0.000[0.001]	0.003	0.155[large]

Table A.10: Relative taxa abundance comparisons at the genus level between the COPD-HIV-, COPD-HIV+, COPD+HIV- and COPD+HIV+ groups in AEC samples. Values displayed as median [interquartile range]. *P-values calculated using the Kruskal-Wallis test; †Effect size calculated based on Kruskal-Wallis H-statistic [Eta squared method]. *Prevotella*∗ - *Prevotella*[f-*Prevotellaceae*], *Prevotella*∓ - *Prevotella*[f-*Paraprevotellaceae*]. Definition of abbreviations: AEC – Airway epithelial cells.

Group1	Group2	Phylum	p-value	†Adjusted p-value
COPD-HIV-	COPD+HIV+	*Fusobacteria*	0.00005	0.0003
		Bacteroidetes	0.00033	0.0019

Table A.11: Pairwise comparisons at the phylum level between the 4 COPD*HIV groups using Dunn's test (only significant comparisons displayed). †Adjusted p-values obtained on applying Bonferroni correction.

Group1	Group2	Genus	p-value	†Adjusted p-value
COPD-HIV-	COPD+HIV+	*Prevotella*∓	0.00001	0.00007
		Prevotella∗	0.00007	0.0004
		Fusobacterium	0.0003	0.002
		Selenomonas	0.0007	0.004
		Neisseria	0.0007	0.004
		Megasphaera	0.001	0.007
COPD-HIV-	COPD+HIV-	*Prevotella*∗	0.0052	0.031
COPD-HIV-	COPD-HIV+	*[Prevotella]*	0.006	0.034
COPD-HIV+	COPD+HIV+	*Paenibacillus*	0.002	0.013

Table A.12: Pairwise comparisons at the genus level between the 4 COPD*HIV groups using Dunn's test (only significant comparisons displayed). †Adjusted p-values obtained on applying Bonferroni correction. *Prevotella*∗ - *Prevotella*[f-*Prevotellaceae*], *Prevotella*∓ - *Prevotella*[f-*Paraprevotellaceae*].

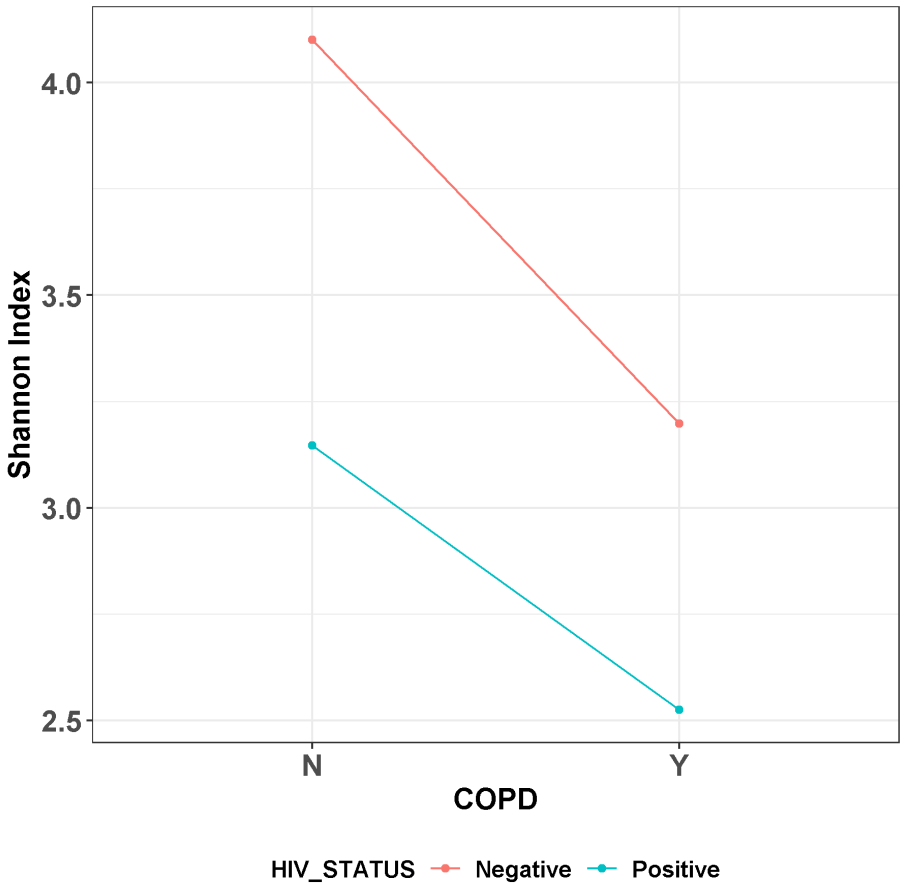
Group 1	Group 2	Sample size	p-value	†q-value
BCW	BW	28	0.002	0.0027
	Brush	101	0.001	0.0015
	CC	49	0.001	0.0015
	EN	29	0.044	0.055
	OWC	50	0.001	0.0015
BW	Brush	79	0.001	0.0015
	CC	27	0.634	0.7315
	EN	7	0.767	0.8218
	OWC	28	0.001	0.0015
Brush	CC	100	0.001	0.0015
	EN	80	0.001	0.0015
	OWC	101	0.001	0.0015
CC	EN	28	0.954	0.954
	OWC	49	0.001	0.001
EN	OWC	29	0.001	0.0015

Table A.13: Pairwise PERMANOVA comparisons between the different specimen types obtained from HIV+ subjects. †q-values obtained using Benjamini-Hochberg FDR correction. Definition of abbreviations: BCW - bronchoscope channel wash; BW - brush water control; CC - cytolyt control; EN - extraction negative; OWC - oral wash control.

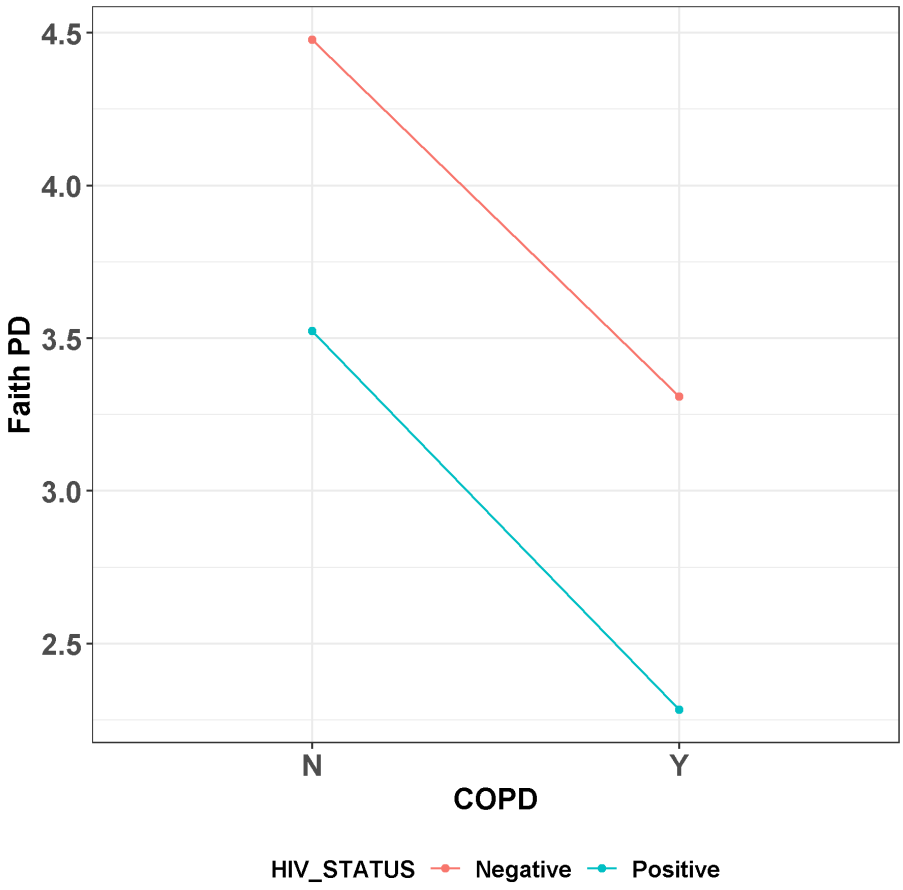
A.1 Interaction Models

We performed a supplementary analysis examining the potential interaction effects between COPD and HIV on the alpha and beta diversity metrics used to characterize the microbiome:

Alpha diversity metrics:
(1) Shannon Index ~ COPD status * HIV status
(2) Faith PD ~ COPD status * HIV status



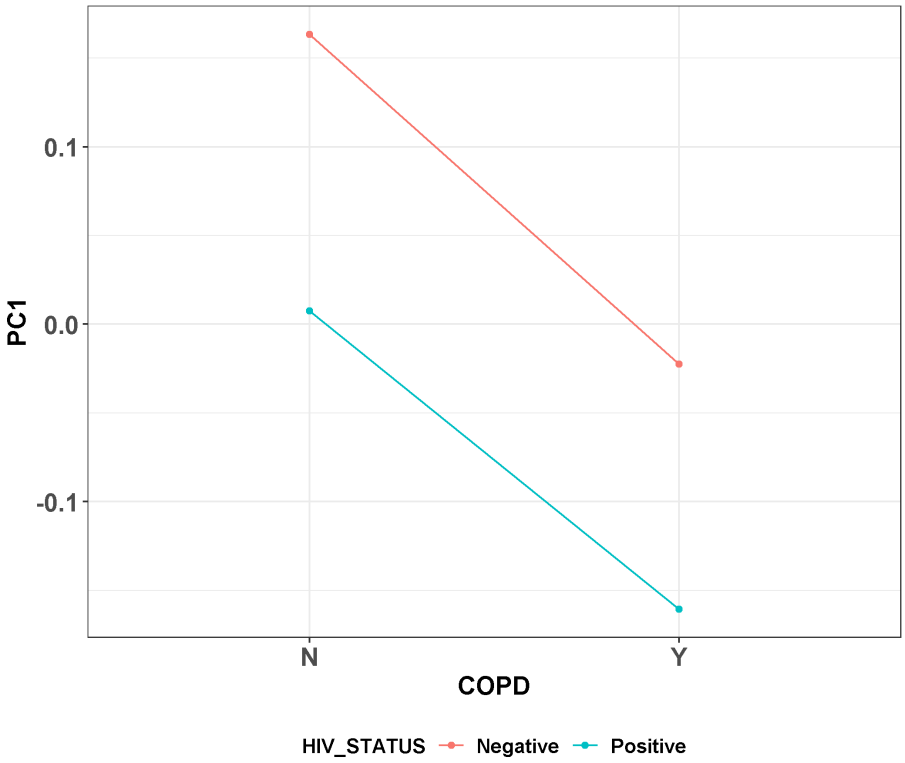
(a)



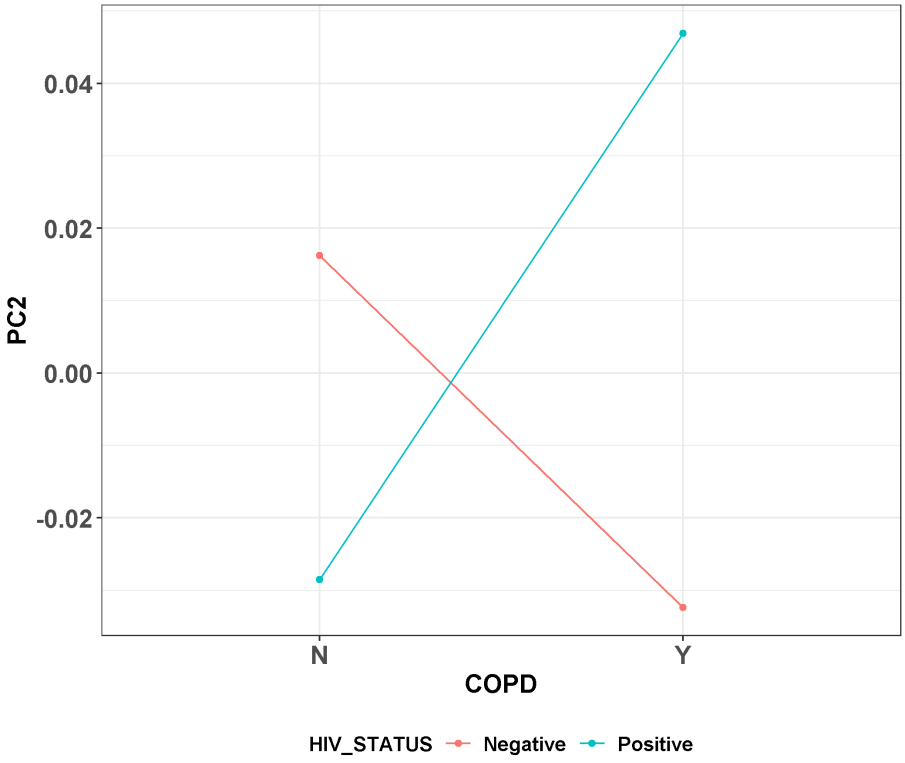
(b)

Figure A.1: Plot showing potential interactive effects of COPD and HIV status on Shannon Index and Faith PD. The lines are basically parallel indicating the absence of an interaction effect. There are no significant P values to report.

Beta diversity metrics:
(1) Bray Curtis PC1 ∼ COPD status * HIV status
(2) Bray Curtis PC2 ∼ COPD status * HIV status



(a)



(b)

Figure A.2: Plot showing potential interactive effects of COPD and HIV status on Bray Curtis PC1 and PC2. The lines in (a) are basically parallel indicating the absence of an interaction effect. The lines in (b) are intersecting indicating a potential interaction effect between COPD and HIV. However, there are no significant P values to report in both cases. This may be because PC1 and PC2 explain only 15.2% and 8.5% of the total observed variation, respectively.

A.2 Decontam Analysis of HIV+ Subjects

A.2.1 Introduction

The R package Decontam [40] was used to identify and visualize contaminating DNA features, and accurately profile sample microbial communities. The interest for *decontam*-ination stems from the fact that contaminating DNA not truly present in the sampled community may be introduced through various external (such as research subjects' or investigators' bodies, sample collection instruments and laboratory reagents, surfaces and air) [1] [147] and internal (during sample mixing, processing or sequencing) sources [110]. Despite employing best laboratory practices, not all contaminating features are eliminated from downstream analysis [190].

In silico methods are also commonly employed to remove contaminants. In our own analysis, we have incorporated some of these methods, however they come with their own limitations - (1) removing DNA features having relative abundance below a certain threshold value - implies risk of losing rare features truly present in the sample, (2) removing features already known as contaminants based on data from previous studies or biological relevance - true sequences may be lost by eliminating features non-specific to the study, and (3) removing features that are present in negative controls - can be significantly influenced by cross-contamination [40].

In this analysis, we use Decontam, which primarily uses two contaminant identification methods: (1) "frequency"-based, where the frequency of each microbial feature is analyzed in relation to the input DNA concentration to identify contaminants, and (2) "prevalence"-based, where the presence/absence or "prevalence" of each ASV in true positive samples is compared to its presence/absence in negative controls to identify contaminants.

A.2.2 Methods

Different sample types including bronchial brushings, bronchoscope channel washes and oral wash controls were obtained for all HIV+ subjects. Other control specimens consisting of brush water controls, cytolyt controls, extraction negatives and no-template controls were also obtained. The microbiome in these different sample types was measured using the 16s rRNA sequencing method and analysed using QIIME 2™. However unlike the main analysis, no samples or ASVs were filtered here (no-template controls were dropped from the analysis when removing mitochondrial and chloroplast DNA).

In our analysis, the "prevalence" contaminant identification method was used, as it is preferred for low-biomass samples. Low-biomass samples are those with low densities of bacterial cells and therefore low quantities of bacterial DNA [8].

A.2.3 Results

The final Decontam analysis included 118 samples of varying specimen types and 3,732 microbial features. Fig. (A.3) indicates the library size (i.e. the number of reads) of each sample, as a function of whether that sample was a true positive sample (i.e. Brush) or a control (i.e. bronchoscope channel wash, brush water control, cytolyt control, oral wash control, and extraction negative). The library sizes of the true positive samples (brush samples) primarily fall between 5,000 to 30,000 reads. Almost all of the control samples have library sizes greater than the true positive samples, but there are a couple of low-read outliers. Contrary to what would be expected, the true positive samples have fewer reads when compared to the controls. This may be because of the low-biomass quality of AEC brushings, which leads to a lower signal to noise ratio, making it is harder capture microbial signals.

The "prevalence" method identified 137 potential contaminants out of a total 3,732 sequence features at the default threshold of 0.1. Fig. (A.4) is a plot shows how the contaminants are able to classify samples into controls and true positive samples, based on the number of times a subset of these taxa were observed in each group. The samples split pretty cleanly, although by a narrow margin, into a branch that shows up mostly in true positive samples, and another that shows up mostly in the controls.

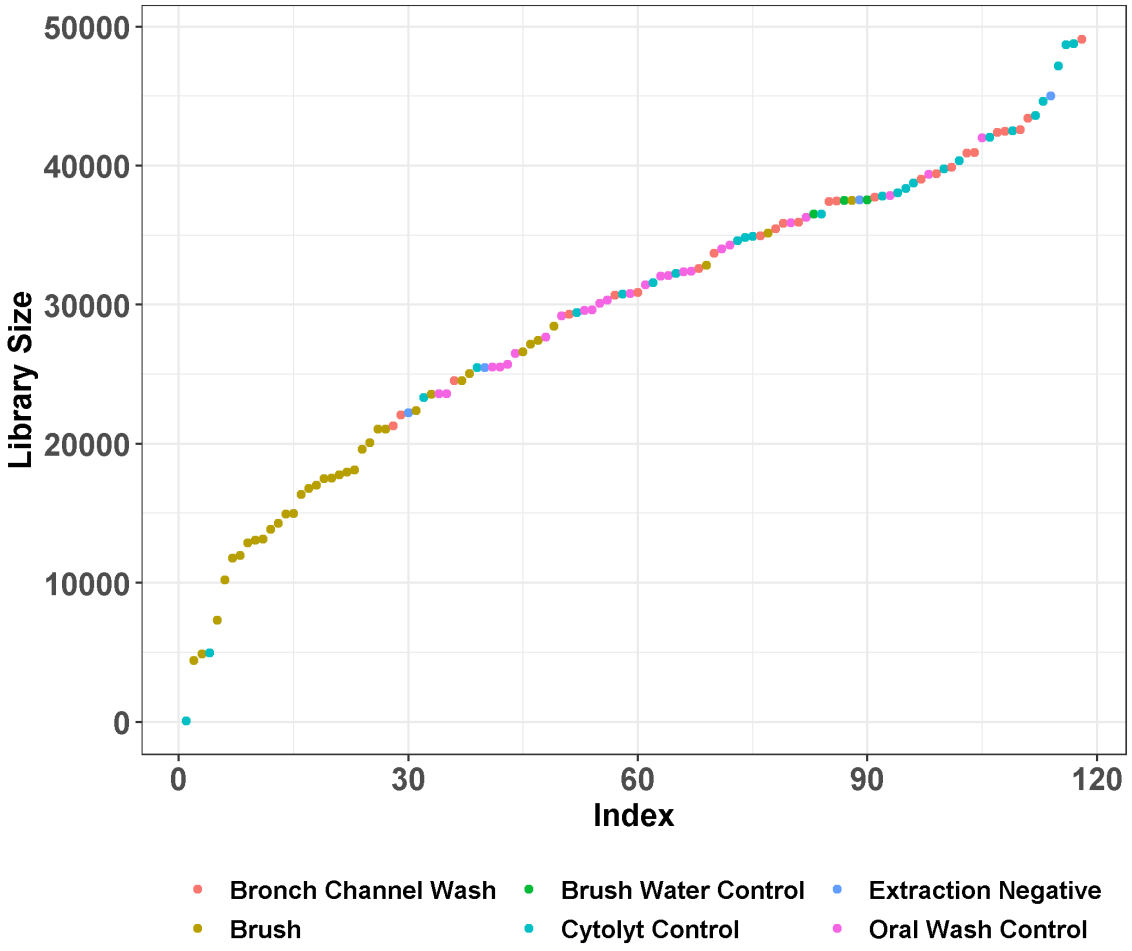


Figure A.3: Number of 16s RNA gene copies/µL observed in control specimens.

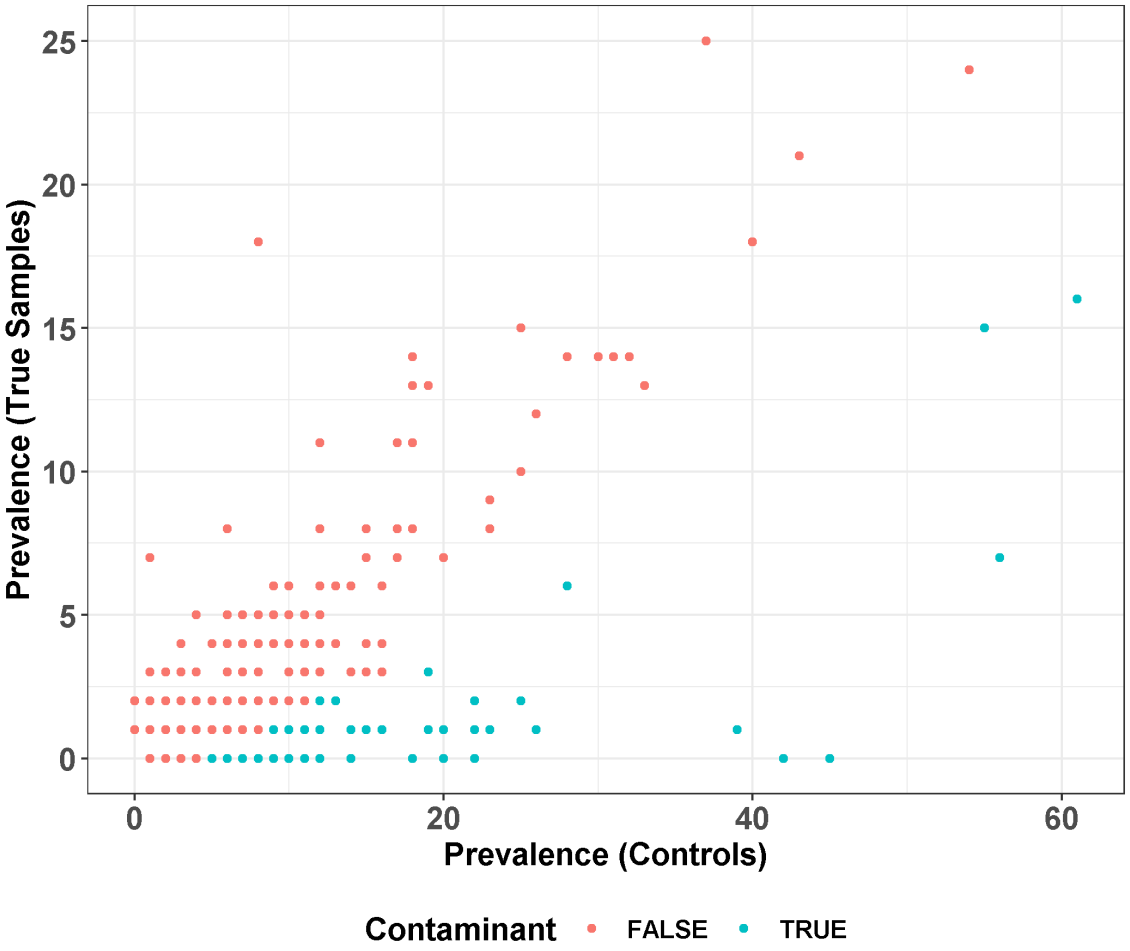


Figure A.4: Frequency of occurrence of taxa in negative controls (consisting of bronchoscope channel washes, brush water controls, cytolyt controls, extraction negatives, and oral wash controls) and true positive samples (brushes).

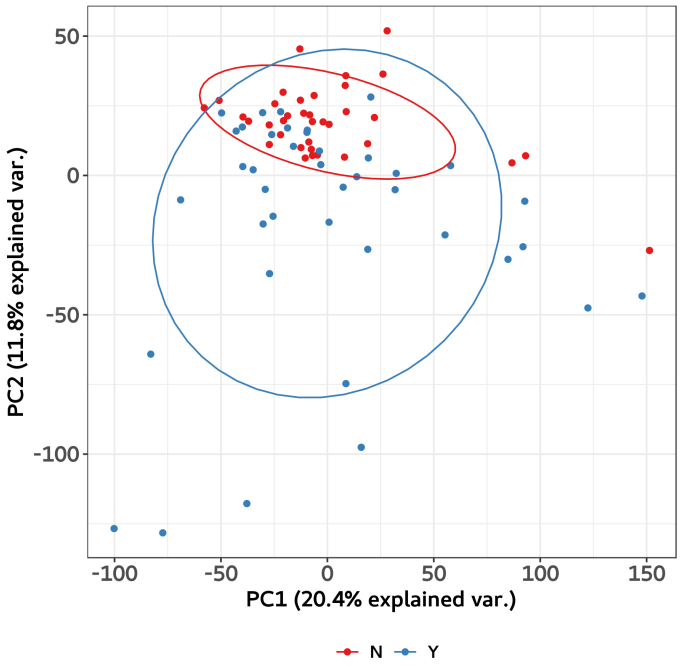
The top 5 contaminants identified are given in Table (A.14).

ASV	∗p-value
Proteobacteria Sphingomonas	2.587e-08
Firmicutes Alicyclobacillus	1.360e-07
Proteobacteria Methylobacterium	1.318e-06
Actinobacteria Candidatus Aquiluna	3.081e-06
Bacteroidetes	6.715e-04

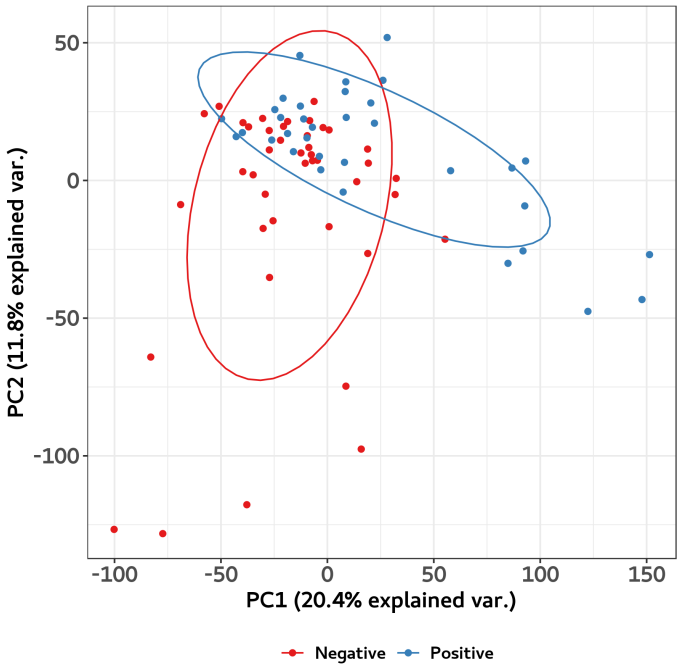
Table A.14: Table showing the top 5 contaminants identified using the "prevalence" Decontam method. ∗p-value obtained from Fisher's exact test. Abbreviations: ASV - Amplicon Sequence Variants.

Appendix B

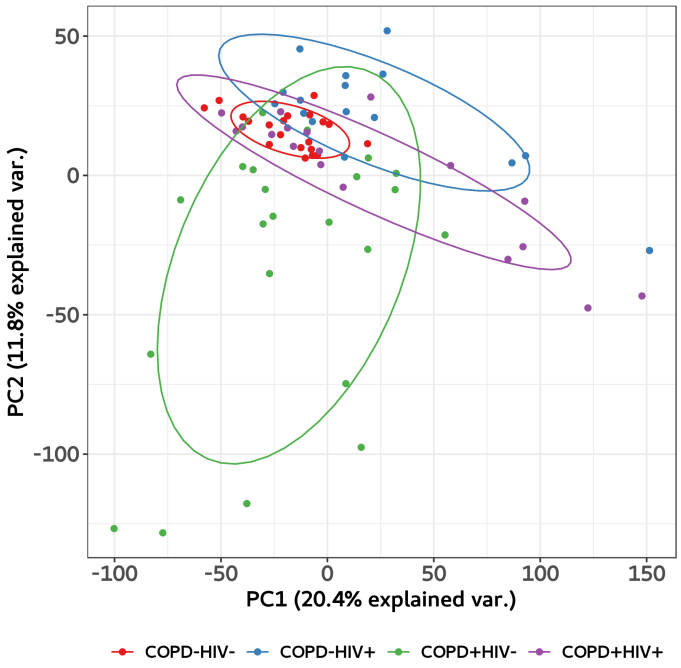
Methylome Analysis


50
0
-50
-100
PC2 (11.8% explained var.)
-100
-50
0
50
100
150
PC1 (20.4% explained var.)
N
Y

(a)


50
0
-50
-100
PC2 (11.8% explained var.)
-100
-50
0
50
100
150
PC1 (20.4% explained var.)
Negative
Positive

(b)

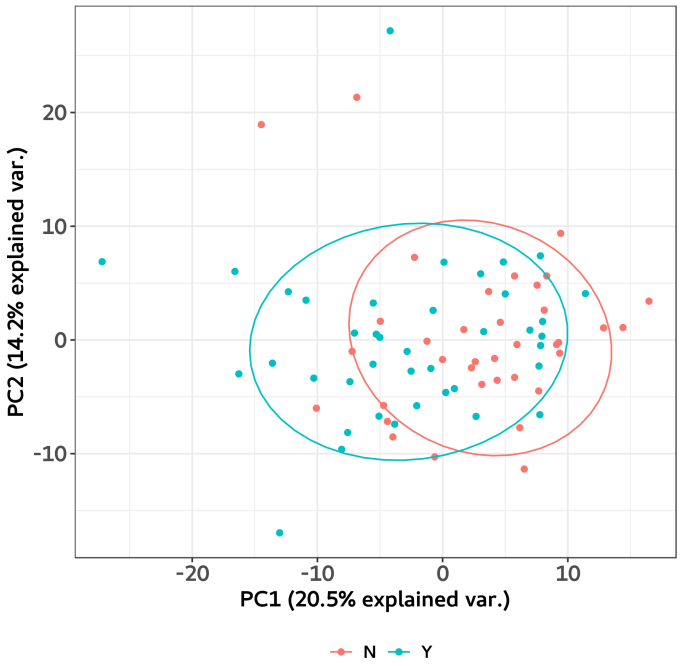


(c)

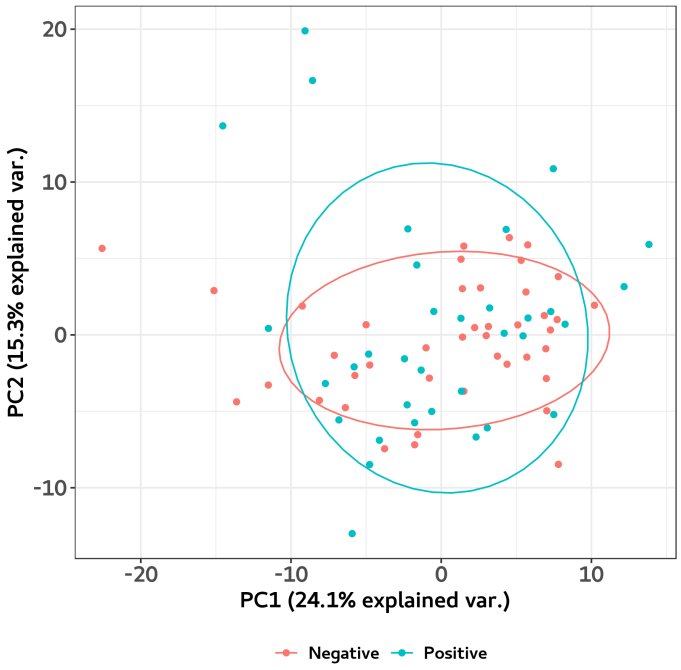
Figure B.1: Principal components analysis of methylation profiles according to (a) COPD, (b) HIV and (c) combined COPD*HIV status. (a) COPD status (COPD- patients (N) – red points; COPD+ patients (Y) – blue points), (b) HIV status (HIV- patients (Negative) – red points; HIV+ patients (Positive) – blue points), and (c) combined COPD*HIV status (COPD+HIV+ – purple points; COPD+HIV- – blue points; COPD-HIV+ – green points; COPD-HIV- – red points); the ellipses enclosing each group are also shown. Definition of abbreviations: PC - principal component.

Appendix C

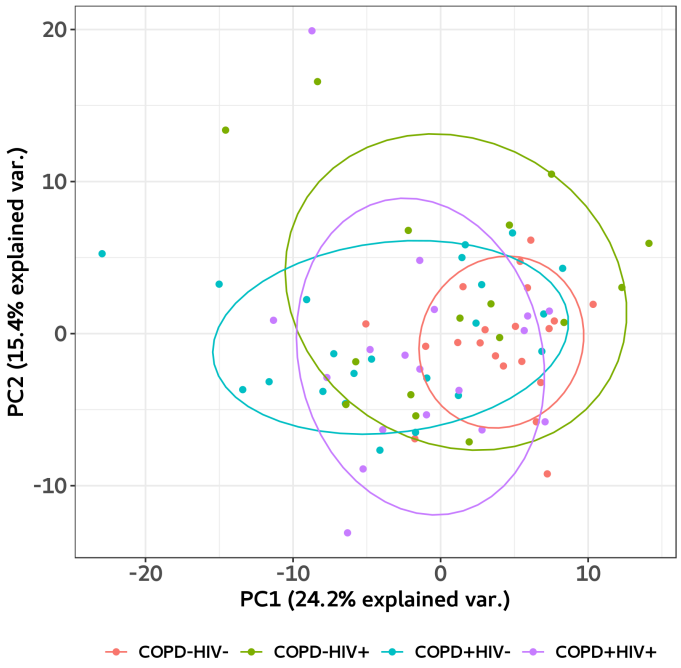
Transcriptome Analysis


20
10
0
-10
PC2 (14.2% explained var.)
-20
-10
0
10
PC1 (20.5% explained var.)
N
Y

(a)


20
10
0
-10
PC2 (15.3% explained var.)
-20
-10
0
10
PC1 (24.1% explained var.)
Negative
Positive

(b)



(c)

Figure C.1: Principal components analysis of gene expression profiles according to (a) COPD, (b) HIV and (c) combined COPD*HIV status. (a) COPD status (COPD- patients (N) – red points; COPD+ patients (Y) – blue points), (b) HIV status (HIV- patients (Negative) – red points; HIV+ patients (Positive) – blue points), and (c) combined COPD*HIV status (COPD+HIV+ – purple points; COPD+HIV- – blue points; COPD-HIV+ – green points; COPD-HIV- – red points); the ellipses enclosing each group are also shown).

Appendix D

Integration Analysis

For the combined COPD*HIV groups, the two-ome integration analyses identified top features between the (a) microbiome and methylome, and (b) microbiome and transcriptome (Fig. D.1). The top three ASV-Gene and ASV-CpG pairs and their respective correlation values are shown in Table (D.1).

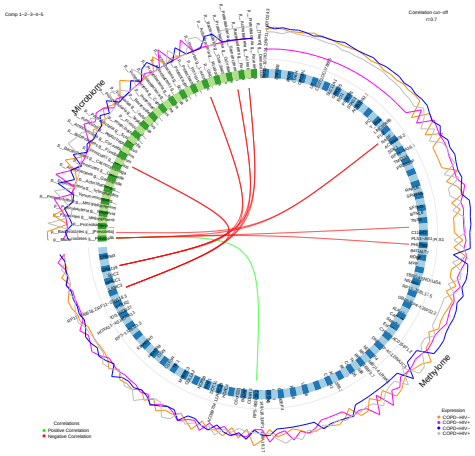
Top ASV-CpG Pairs	Correlation
p_Bacteroidetes g_Prevotella - CpG FUZ	-0.774
p_Bacteroidetes - CpG GPR139	-0.763
p_Bacteroidetes g_Fluviicola - CpG GPR139	-0.762

(a)

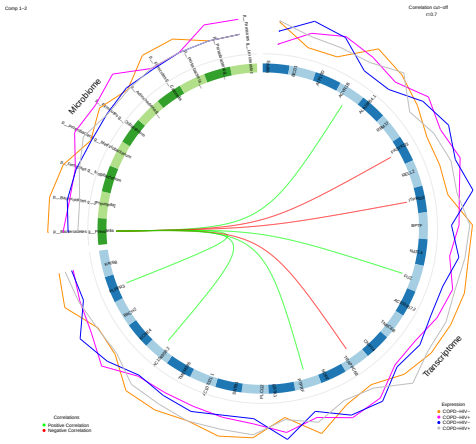
Top ASV-Gene Pairs	Correlation
p_Bacteroidetes g_Prevotella - FUZ	0.765
p_Bacteroidetes g_Prevotella - FASTKD3	-0.742
p_Bacteroidetes g_Prevotella - TRAPPC6B	-0.733

(b)

Table D.1: Table showing top (a) ASV-CpG (microbiome + methylome) and (b) ASV-Gene (microbiome + transcriptome) pairs obtained and their respective correlation values corresponding to the combined COPD*HIV effect. (Minus sign indicates negative correlation)



(a)



(b)

Figure D.1: DIABLO circos plot showing the within and between correlations between the (a) microbiome and methylome, and (b) microbiome and transcriptome. The three -omes are represented on the side quadrants; expression levels of each variable according to combined COPD*HIV status can be viewed along the circumference. Positive correlation - Green ; Negative correlation - Red; Correlation cutoff = 0.7.