IMPERIAL

Analysis of microbiome-based asthma phenotypes

Ali Versi

Imperial College London (NHLI & DSI)

This thesis is submitted for the degree of Doctor of Philosophy (PhD)

Abstract (300 words)

Background: Asthma encompasses several clinical phenotypes. Metagenomic sequencing enables species-level characterisation and functional profiling. Characterisation of the airway microbiome, dysbiosis phenotypes and host-microbial crosstalk is wanting.

Methods: Metagenomic sequencing of induced sputum from asthmatic patients (n=148) was performed. Patient data was stratified by asthma severity, inflammatory status and human transcriptome-associated clusters (TACs). Dysbiosis phenotypes associated with high abundance bacterial species were determined. A relative dominant species (RDS) was defined and characterised using sputum transcriptomics. Crosstalk between host immune responses and the microbiome was evaluated in early- (EO) and late-onset (LO) asthma.

Results: Reduced alpha-diversity occurred with severity and was accompanied by an increase in *Haemophilus influenzae* (Hi) and *Moraxella catarrhalis* (Mc) in non-smokers. Hi and *Tropheryma whipplei* (Tw) were increased in ex- and current smokers.

53% of samples were classified as RDSs with Hi RDS being most common (31%). Hi RDS had the longest disease duration. Clustering of RDSs revealed a cluster (n=9) of exclusively Hi RDSs with a longer duration of disease, higher sputum neutrophilia associated with enrichment of MAPK, NF-κB, TNF, mTOR and necroptosis pathways. This cluster was also enriched for neutrophil extracellular traps (NETosis) and Th-17 activation.

Host-microbial integrative analysis revealed shared pathways between EO and LO phenotypes although the constituent genes and species were different. Barrier function and adhesion molecule-associated pathways were associated with EO asthma. Mc was only associated with pathways in EO and Tw with pathways only in LO.

Conclusions: Hi and Mc were more abundant in neutrophilic asthma and linked to inflammasome. Hi and Tw were linked to TAC1 and ILC2 signatures. A Hi abundant cluster with neutrophilic activation and NETosis driven by host-microbial interaction was elucidated in EO asthma. Shared pathways with unique host-microbe interactions exist between EO and LO asthma indicating distinct mechanisms underpinning the heterogeneity of asthma immunopathophysiology.

Acknowledgements

I am grateful to have had this opportunity to pursue a PhD at Imperial College London would like to express my sincere gratitude to all those who have supported me throughout my doctoral journey.

First and foremost, I would like to thank my supervisors Professor Ian Adcock, Professor Kian Fan Chung and Dr Nazanin Kermani. Throughout the course of my PhD Professor Adcock has pushed me to develop my research skills, provided support and flexibility during my numerous health complications and been a wonderful mentor. I am especially grateful for his tireless effort in delivering rapid feedback to my work and his patience with some of my early drafts. I have immensely enjoyed working with and being mentored by Professor Fan Chung, I have grown as a researcher tremendously learning to write papers with him and have been inspired by his dedication to research. I would like to thank Dr Nazanin Kermani, who has provided support from the department of computing and the data science institute. Her ability to easily explain complex concepts has enabled me to grasp many concepts quickly in the realm of bioinformatics, data analysis, machine learning and AI.

I would like to express my gratitude to collaborators I have worked with during my PhD. I am very grateful to have had the opportunity to collaborate with Dr Adnan Azim. His mentorship and expertise have profoundly enriched my PhD journey. I would like to thank Dr Fransiskus Xaverius Ivan, whose mentorship helped me develop my practical bioinformatics skills and Dr Mahmoud Ibrahim who has always been providing very insightful feedback. I would also like to thank their respective PIs, Professor Peter Howarth, Professor Sanjay Chotirmall and Professor Anke-Hilse Maitland-van der Zee who have all leant me their precious time and insights.

I am also extremely grateful for the collaborations I had with GSK and the support the team provided me with. In particular, I would like to thank John Riley, Dr Andy Nuzzo and Dr Stewart Bates. I would like to thank Dr Pankaj Bhavsar and the wider NHLI community including Dr Katie Raby and the wider community at the DSI.

I would also like to thank Professor Sun Kim and The Bio & Health informatics AI research group at Seoul National University for hosting my Turing Fellowship.

I would like to thank Dr Yusef Badi and Dr Hisham Abubakar-Waziri for being great colleagues and wonderful friends. In addition, I would like to thank Tony Umleo and Shuet Kwan Tang for assisting me with the bureaucracy during my health issues.

Lastly, I would like to thank my parents, whose sacrifices have enabled me to take this PhD journey.

Statement of Originality

The contents of this thesis are my own original work and work besides my own has been appropriately referenced.

Copyright Declaration

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

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

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

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

Table of contents:

Abstract (300 words).....2

Acknowledgements.....4

Statement of Originality.....6

Copyright Declaration.....7

Table of contents:.....8

List of abbreviations:.....13

List of Tables:.....16

List of Figures:.....17

Chapter 1. Introduction & Background.....20

1.1 Asthma.....20

1.2 Epidemiology.....23

1.3 The pathophysiology of asthma.....25

1.4 The Airway Microbiome.....28

1.4.1 The airway microbiome.....28

1.4.2 Lung microbiome in respiratory diseases: Focus on asthma.....30

1.5 Asthma Treatment: current approaches and the respiratory microbiome.....32

1.5.1 Conventional asthma treatment.....32

1.5.2 The lung microbiome and asthma treatment.....35

1.6 Phenotypes and endotypes: A place for molecular phenotyping in asthma. 36

1.6.1 Asthma phenotypes:.....36

1.6.2 Asthma endotypes.....37

1.6.3 Molecular phenotyping: advancements and gaps.....37

1.7 Towards bridging clinical gaps: Precision medicine in asthma.....41
1.8 Thesis Aims.....44
Chapter 2. Datasets & Methodology.....45
2.1 Introduction.....45
2.2 Datasets.....45
2.2.1 U-BIOPRED.....45
2.2.2 Participants.....45
2.2.3 Samples collected.....47
2.2.4 Sputum induction and initial sample handling.....47
2.2.5 Statistical Methods.....51
2.2.6 Differentially expressed gene (DEG) analysis.....52
2.2.7 Pathway Enrichment.....53
2.2.8 Derivation of transcriptomic-associated clusters (TACs).....54
Chapter 3. Evaluation of metagenomic pipelines for U-BIOPRED metagenomic sequence data.....56
3.1 Introduction.....56
3.1.1 Shotgun metagenomics.....56
3.1.2 Bioinformatics pipeline for taxonomy classification and functional genomics.....57
3.1.3 Kraken, Kaiju and MetaPhlAn: Three metagenomic pipelines.....57
3.2 Methods.....60
3.2.1 Participants:.....60
3.2.2 Sample collection, Library preparation and Sequencing:.....60
3.2.3 Quality control and host reads removal:.....61
3.2.4 Bioinformatics Pipelines:.....61

3.3 Results.....62

3.3.1 Microbiome profiling with MetaPhlAn2.....64

3.4 Discussion.....69

Chapter 4. Haemophilus influenzae and Moraxella catarrhalis in the sputum of severe asthma associated with inflammasome and neutrophil activation.....71

4.1 Introduction.....71

4.2 Methods.....72

4.2.1 Quality control of metagenomic data and host reads removal.....72

4.2.2 Statistical and computational analysis.....73

4.3 Results.....73

4.3.1 α- and β-diversity.....81

4.3.2 Differentially abundant genera and species by asthma grouping.....86

4.3.3 Species abundance and clinical and inflammatory features.....89

4.3.4 Species abundance according to granulocytic inflammation.....91

4.3.5 Species abundance according to TAC molecular phenotypes.....91

4.3.6 Longitudinal stability of microbiome according to asthma severity.....92

4.4 Discussion.....95

Chapter 5. A severe neutrophilic asthma phenotype with high airway Haemophilus influenzae abundance.....101

5.1 Introduction.....101

5.2 Method.....102

5.2.1 Participants.....102

5.2.2 Sample collection, DNA/RNA extraction and data pre-processing.....104

5.2.3 Definition of relative dominant species (RDSs).....104

5.2.1 Method for designating a relative dominant species (RDS).....106

5.2.2 Clustering on abundance of low-diversity bacterial species.....109
5.2.1 Determiniation of dominant RDS.....109
5.2.2 Longitudinal analysis.....110
5.2.3 Statistical analysis.....110
5.3 Results.....110
5.3.1 Low a-diversity RDSs.....110
5.3.2 Haemophilus influenzae, Moraxella catarrhalis and Tropheryma whipplei RDSs 119
5.3.3 Cluster analysis of low diversity samples.....126
5.3.4 Haemophilus influenzae RDSs of C1 and C2 clusters.....133
5.3.5 Longitudinal Analysis of severe asthma samples.....143
5.4 Discussion.....144
Chapter 6. Host-microbial interactions differ with age of asthma onset.....148
6.1 Introduction.....148
6.2 Method.....149
6.2.1 Participants.....149
6.2.2 Induction and initial sample handling.....149
6.2.3 Statistical and computational analyses.....150
6.2.4 Hyperparameter tunning.....151
6.2.5 Sparse CCA correlation.....151
6.2.6 Pathway enrichment of genes.....151
6.2.7 Visualization of shared host pathways and disease-specific host pathways.....152
6.3 Results.....153
6.4 Discussion.....161

Chapter 7. Summary & Discussion.....164
7.1 Summary.....164
7.1.1 Data processing and pipeline evaluation.....164
7.1.2 Microbiome evaluation by phenotype.....165
7.1.3 Relative dominant species based phenotypes.....168
7.1.4 Microbiome-host gene expression interaction analysis.....169
7.2 Discussion.....170
7.2.1 Impact on Landscape.....170
7.2.2 Limitations.....175
7.2.3 Future Studies.....177
Chapter 8. References.....179
Chapter 9. Appendix.....217
Appendix I. Gene signatures used in this thesis.....218
Appendix II. Permissions for reproduction of materials.....223

List of abbreviations:

ACQ5	Asthma Control Questionnaire
ANOVA	Analysis of Variance
BL	Baseline Timepoint
bp	Base Pair
CCA	Canonical Correlation Analysis
CDF	Cumulative Distribution Function
COPD	Chronic Obstructive Pulmonary Disease
DAMP	Damage-Associated Molecular Pattern
DEG	Differentially Expressed Gene
DNA	Deoxyribonucleic Acid
EO	Early-onset
ES	Enrichment Score
FDR	False Discovery Rate
FeNO	Fractional Exhaled Nitric Oxide
FEV_1	Forced Expiratory Volume in The First Second
GC	Guanine-Cytosine
GEM	Genome Scale Metabolic Model
GINA	Global initiative for Asthma
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GSVA	Gene Set Variation Analysis
HC	Healthy Non-Smoking Controls
ICS	Inhaled Corticosteroid
IL	Interleukin

IQR	Interquartile Range
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LABA	Long-Acting Beta-Agonist
LAMA	Long-Acting Muscarinic Agents
LO	Late-onset
LT	Longitudinal Timepoint
LTRA	Leukotriene Receptor Antagonist
MMA	Mild-Moderate Asthmatic Non-Smokers
NET	Neutrophil Extracellular Trap
NGS	Next-Generation Sequencing
OCS	Oral Corticosteroid
PAMP	Pathogen-Associated Molecular Pattern
PCoA	Principle Coordinate Analysis
PCR	Polymerase Chain Reaction
PERMANOVA	Permutational Multivariate Analysis of Variance
PID	Pathway Interaction Database
RDS	Relative Dominant Species
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
SA	Severe Asthmatic
SABA	Short-Acting Beta-Agonist
SAn	Severe Asthmatic Non-Smoking
Sas/ex	Severe Asthmatic Smokers and Ex-Smokers
TAC	Transcriptome-Associated Clusters
Th	T helper

TLR	Toll-like Receptor
TSLP	Thymic Stromal Lymphopoietin
U-BIOPRED	Unbiased BIOmarkers in PREDiction of respiratory

List of Tables:

Table 1. Criteria for participants of each adult cohort in the U-BIOPRED study.....46

Table 2. Summary of the output of the three pipelines.....63

Table 3. Characteristics of participants.....75

Table 4. Characteristics of subjects classified by sputum granulocytic inflammation type.....78

Table 5. Characteristics of subjects classified by Transcriptome-Associated Cluster (TAC).....80

Table 6. Characteristics of relative dominant species (RDS) and non-RDSs.....111

Table 7. Presence and abundance of bacterial species in sputum of severe asthma. 115

Table 8. Characteristics of severe asthma with Haemophilus influenzae (Hi), Moraxella catarrhalis (Mc) & Tropheryma whipplei (Tw) relative dominant species (RDS).....122

Table 9. Abundance of RDS species present in C1 and C2 clusters and in MMA/HC 128

Table 10. Characteristics of clusters C1 and C2, and non-RDSs.....131

Table 11. Characteristics of severe patients with the Haemophilus influenzae RDSs samples in C1 and C2 clusters.....134

Table 12. Characteristics of Adult onset and Child onset.....154

List of Figures:

Figure 1 Mast cell degranulation triggers airway inflammation (18).....21

Figure 2. T2-high endotype (19).....22

Figure 3. Deaths and disability due to asthma 1990–2019 (24).....24

Figure 4. Airway remodelling in Asthma (31).....25

Figure 5. Pathophysiology of T2-high and T2-low asthma (33).....26

Figure 6. Clinical phenotypes of asthma as T2-high and T2-low endotypes (39).....27

Figure 7. Signals from the respiratory microbiome influence lung immunity (67).....30

Figure 8. Global Initiative for Asthma treatment guidelines (93).....34

Figure 9. Multiple omics data types and approaches to disease research (126).....38

Figure 10. Consort Diagram of Adult Cohorts with sputum samples collected and the timepoints of collection in U-BIOPRED.....48

Figure 11. Venn Diagram depicting data overlap of available platforms in the sputum 51

Figure 12. TAC definition.....55

Figure 13. Cohort information after metagenomic sequencing of sputum.....62

Figure 14. Assessment of metagenomic sequencing and overview of phylum-level microbiome profiles.....64

Figure 15. Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes across sample batches B1-B5.....66

Figure 16. The Venn diagram depicts the overlap for the two datasets at each taxa-level.....68

Figure 17. Airway of severe asthma subjects exhibits distinct species-level microbiome profiles.....81

Figure 18. Heatmap of Spearman's correlation coefficients.....82

Figure 19. Sputum inflammation type shows distinct species-level microbiome profiles.....84

Figure 20. Molecular phenotypes of asthma exhibit distinct microbiome profiles.....85

Figure 21. Airway of severe asthma subjects exhibits distinct genus-level microbiome profiles.....86

Figure 22. Barplots of the microbiome composition by smoking status.....87

Figure 23. Barplots of the microbiome composition by OCS usage.....88

Figure 24. Heatmap of Spearman's correlation coefficients by species and severity89

Figure 25. Airway of severe asthma subjects exhibits temporal stability in species-level microbiome profiles.....93

Figure 26. Airway of severe asthma subjects exhibits temporal stability in genus-level microbiome profiles.....94

Figure 27. Consort diagram depicting subjects in analysis.....103

Figure 28. Flowchart showing the determination of relative dominant species (RDS) by α-diversity threshold in the sputum samples and applied to samples of severe asthma patients and hierarchical clustering of RDS samples.....104

Figure 29. Relative dominant specie (RDS) cut-off methodology.....106

Figure 30. Distribution plot of alpha-diversity for MMA and HC Cohorts.....108

Figure 31. Presence and abundance of bacterial species by phylum.....114

Figure 32. Relative dominant specie (RDS) co-occurrence heatmap.....118

Figure 33. Transcriptomic differentially-expressed genes between RDS groups and MMA/HC cohort.....120

Figure 34. GO enrichment dot-plots between the Haemophilus influenzae (panel A) and Moraxella catarrhalis relative dominant species (panel B) compared to MMA group.....125

Figure 35. Metagenomic hierarchical clustering of sputum metagenomic species-level abundance of relative dominant specie samples (k=2).....126

Figure 36. Diversity of Aitcheson distance clusters of RDS samples and Bacterial Species composition.....127

Figure 37. Composition log-fold change between each of the C-Hi (A) and C2-Hi (B) clusters and the MMA/HC group using ANCOM-BC.....128

Figure 38. DEG and pathway analysis of clusters C1 vs C2.....133

Figure 39. Haemophilus influenzae component of C1 (C1-Hi) and C2 (C2-Hi) clusters of RDS samples.....137

Figure 40. Gene set variation analysis (GSVA) enrichment score between the C1 and C2 cluster component of Haemophilus Influenzae RDS (C1-Hi & C2-Hi) and the Non-Hi-RDS.....138

Figure 41. Comparison of C1-Hi, C2-Hi and Non-Hi-RDS groups.....139

Figure 42. Microbial pathway prevalence in C1-Hi and C2-Hi cluster samples.....141

Figure 43. ANCOM-BC plot depicting differentially prevalent microbial pathways between C2-Hi and C1-Hi.....142

Figure 44. Sankey plot of the stability of dominant RDS over time.....143

Figure 45. Flow Chart showing the overall method of sample collection, cohort grouping and the application of sparse CCA.....153

Figure 46. GSVA sample-wise enrichment score correlation with clinical characteristics for each component.....156

Figure 47. Component-genes pathway analysis.....157

Figure 48. Visualization of Leukocyte Transendothelial Migration and Focal Adhesion for both LO and EO asthma and GSVA sample-wise enrichment score correlation with clinical characteristics.....159

Figure 49. Pathway analysis by dominant species in components for each onset-group.....160

Chapter 1. Introduction & Background

1.1 Asthma

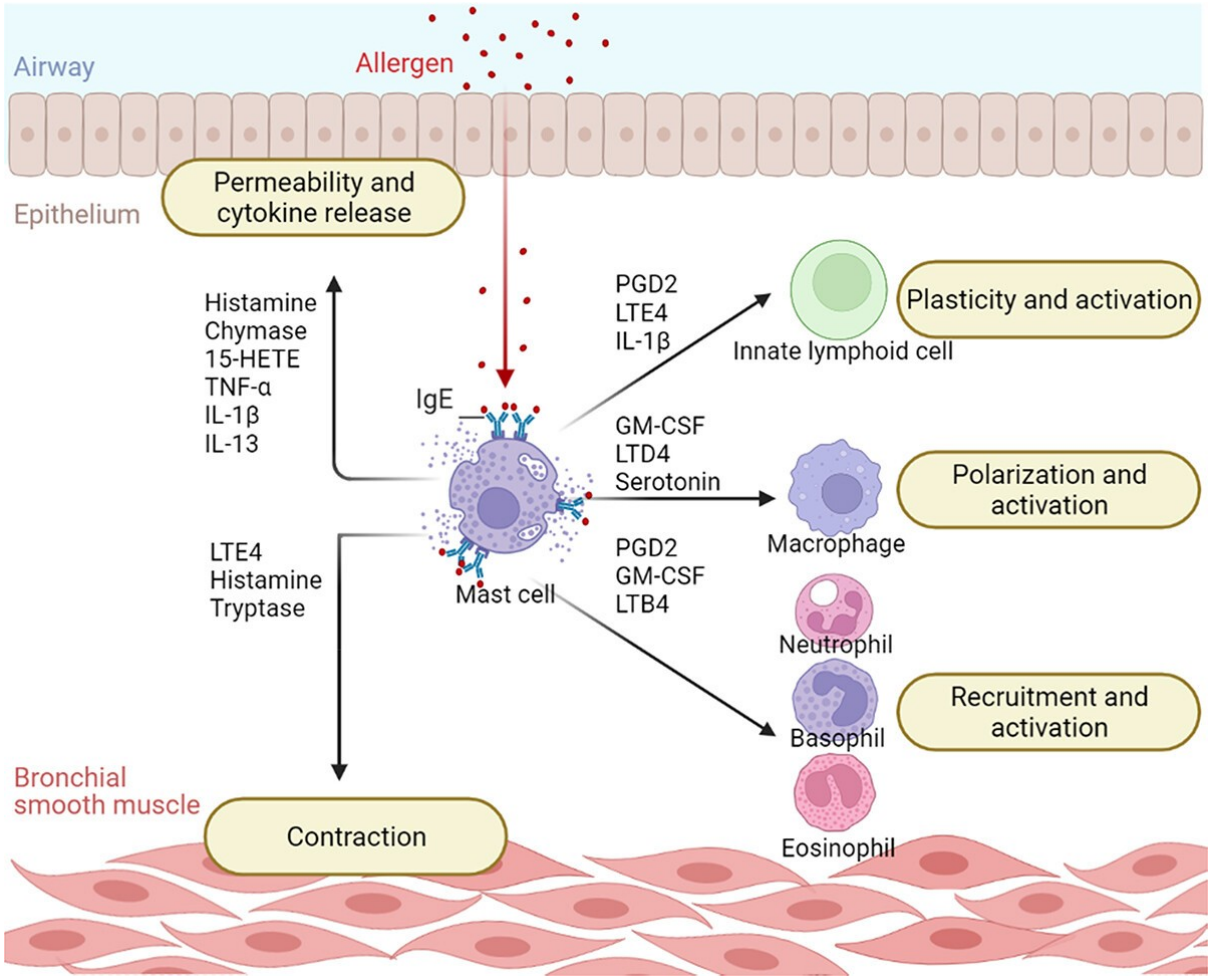
Asthma, one of the most common diseases worldwide, is defined by the Global initiative for Asthma (GINA) as a heterogeneous disease characterised by chronic inflammation, recurring episodes of wheezing, breathlessness, chest tightness and coughing (1). It is estimated that more than 300 million people worldwide of all ages suffer from asthma and it is the most common type of airway disease in children and young adults (2) with its prevalence doubling over the last 20 years (3,4). The United Kingdom has the highest incidence of adult and childhood asthma in Europe (5,6) with an overall prevalence of 6.5% in 2016 (7) and overall annual cost incurred to the United Kingdom public sector estimated to be in excess of £1.1 billion (8). It is estimated that 50% of annual healthcare costs are attributed to the most severe asthmatics (9) those experiencing worse control of asthma symptoms and signs. While the majority of people with asthma respond well to inhaled corticosteroids (ICS) with or without long-acting beta$_2$ agonists (LABA) and have their disease well managed and controlled, some 5-10% of asthmatics have poorly controlled disease (10). These severe asthmatics require high dose ICS in addition to a second controller, such as systemic or oral corticosteroids (OCS) in an attempt to better control the disease better (11). The reliance on the dose of ICS/OCS to define asthma severity adds a major confounding factor to understanding asthma.

Due to the heterogeneous clinical presentation of asthma and the complex interplay between genetic and environmental factors, the focus of the definition of the disease has been aimed at observable characteristics in the clinic rather than the underlaying mechanisms by which the disease manifests (12,13). Therefore, treatment approaches have been shaped by disease severity rather than the fundamental mechanisms involved for each patient (14).

Severe asthma is a chronic inflammatory disease of the airways, with complex and heterogenous clinical presentation (15). Among severe asthma subtypes, severe eosinophilic or type-2 (T2) high asthma is the most established inflammatory phenotype and is characterized by an increased production of type 2 inflammatory

markers, namely interleukin (IL) 4, IL-5 and IL-13 cytokines (16). Mast cell degranulation (or activation) (**Figure 1**) is mainly triggered by allergens interacting with IgE forming an allergen-IgE complex, which in turn binds to the high affinity IgE receptor (FcεRI) on the surface of mast cells (17).

Figure 1 Mast cell degranulation triggers airway inflammation (18)



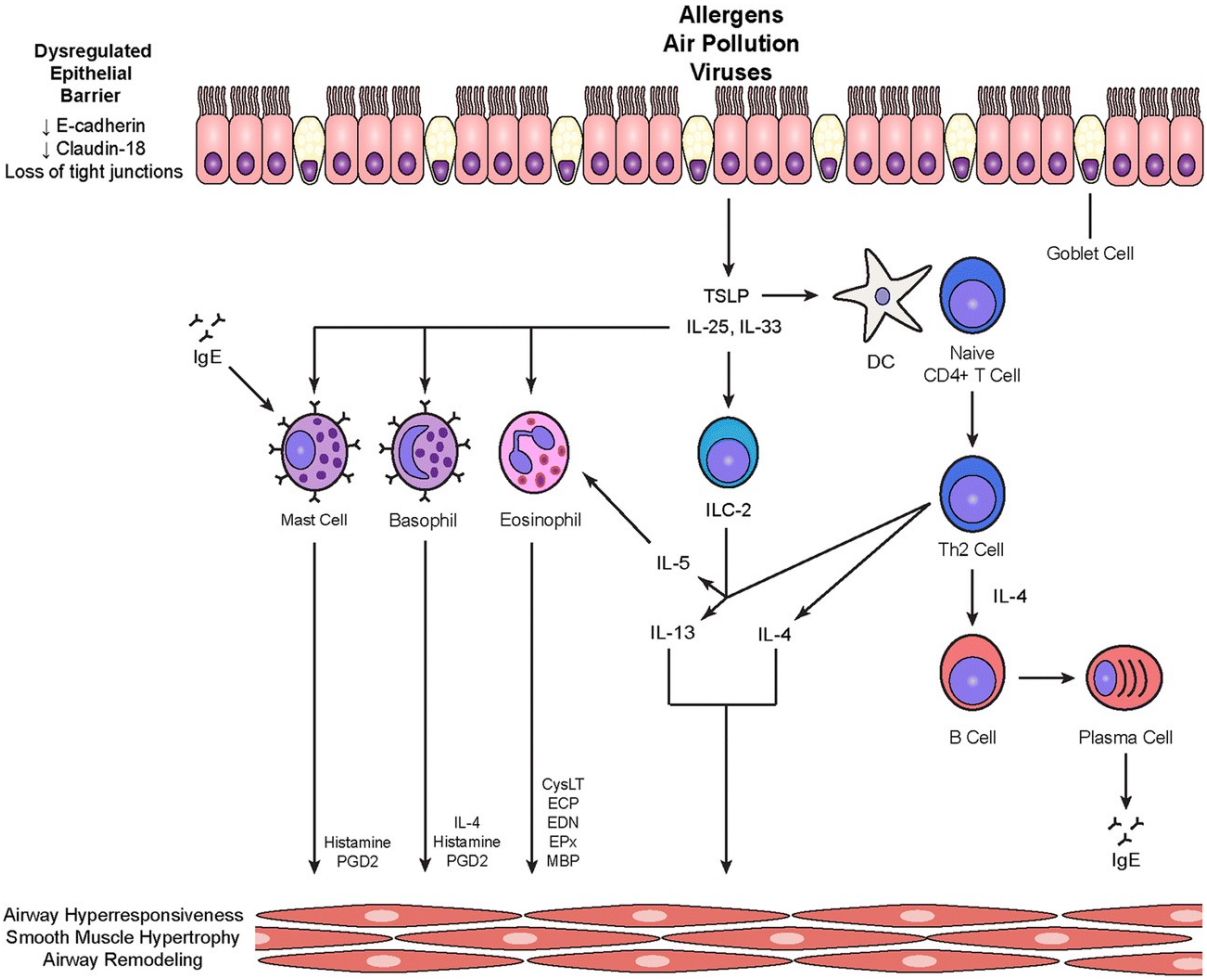
Legend: Mast cells can release a variety of mediators that drive inflammation by activating other cells, triggering bronchoconstriction and altering the permeability of the bronchial barrier.

Reproduced under creative commons license © 2022, Taylor & Francis Group. See appendix: copyright permissions.

As shown in Figure 2, reduced barrier property of the airway epithelial barrier allows allergens and microbes to translocate into the stromal tissues. The airway epithelial cells (AECs) detect these foreign bodies through the formers' pattern recognition receptors (PRRs) and release cytokines called alarmins such as thymic stromal lymphopoietin (TSLP), IL-25, and IL-33. Specificly, TSLP activates dendritic cells

(DC) which would subsequently migrate to the closest lymph nodes and stimulate the differentiation of naive T cells into T-helper cell type 2 (Th2). Th2 cells will then release interleukin (IL) 4, IL-5 and IL-13 cytokines (19).

Figure 2. T2-high endotype (19)



Legend: T2-high inflammatory pathways in asthma. A dysregulated epithelial barrier facilitates translocation of allergens, air pollution, and viruses, leading to release of alarmins such as thymic stromal lymphopoietin (TSLP), IL-25, and IL-33. TSLP primes dendritic cells to induce the differentiation of naïve T cells into Th2 cells. Th2 cells activate B cells via IL-4 to differentiate into plasma cells that generate IgE required for mast cell responses to allergens. The alarmins IL-25 and IL-33 can activate group 2 innate lymphoid cells (ILC2s), mast cells, eosinophils, and basophils. Activated ILC2s, like Th2 cells, produce IL-5 and IL-13. IL-5 promotes eosinophil differentiation and survival. IL-13, IL-4, and inflammatory mediators from mast cells, basophils, and eosinophils have effects on airway hyperresponsiveness, smooth muscle hypertrophy, and airway remodeling. CysLT, cysteinyl leukotrienes; ECP, eosinophil cationic protein; EDN, eosinophil-derived neurotoxin; EPx, eosinophil peroxidase; MBP, major basic protein; PGD2, prostaglandin D2.

Reproduced with permission © 2018, Springer. See appendix: copyright permissions.

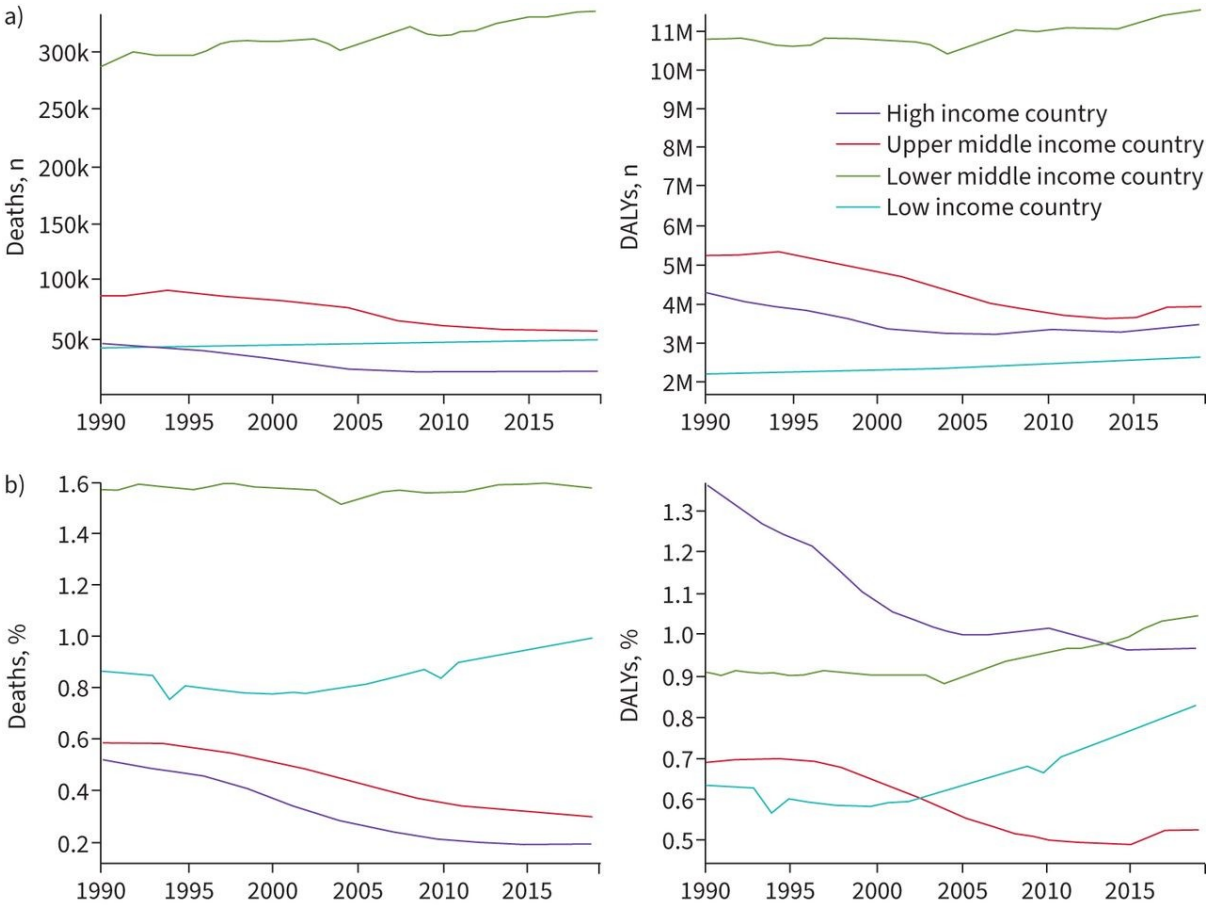
Conversely, type 1 inflammation, activation of the inflammasome activation and subsequent sputum neutrophilia are characteristic of the neutrophilic inflammatory phenotype (20). However, the mechanism(s) underlying neutrophil recruitment and activation have yet to be elucidated, despite evidence linking inflammasome activation and neutrophil airway recruitment in asthma (21). The demonstration of asthma molecular phenotypes is needed to clarify the potential connection between these species and the host immunological and inflammatory responses (20).

1.2 Epidemiology

Epidemiological data reveal the persistence of asthma's global burden despite the decline in age-standardised rates of asthma prevalence, mortality and disability-adjusted life years (DALYs); In the three decades (between 1990 and 2019), the prevalence of asthma decreased by almost 25%, accompanied by a 50% reduction in mortality and 42.55% loss in DALYs probably reflecting the use of ICS (22). More than 262 million prevalent cases of asthma were estimated in 2019, reflecting the disease's continued global health burden (22).

Geographical variations in asthma burden were also clear; areas with high socio demographic index are afflicted with a higher incidence of asthma, while those with a low socio demographic index suffer from higher asthma-related mortality rates (22,23). Western Europe and High-income North America were among the regions with the highest number of prevalent asthma cases, while mortality was higher in lower income areas such as South and Southeast Asia (22). This latter point may reflect the reduced access to inhaled drugs in low- and middle-income countries due to their relatively high costs (24).

Figure 3. Deaths and disability due to asthma 1990–2019 (24)



Legend: a) Number of deaths and disability-adjusted life years (DALYs) due to asthma according to World Bank income category for both sexes and all ages. Between 1990 and 2019 the number of deaths and DALYs in lower middle income countries far exceeded those of other income brackets. b) Percentage of deaths and DALYs due to asthma according to World Bank income category for both sexes and all ages. In 2019, rates of asthma deaths were highest in lower middle income countries, followed by low income countries. By 2019, disease burden (DALYs) in lower middle income countries had overtaken that of high income countries. Both death rates and DALYs have been increasing in low income countries, and DALYs in lower middle income countries. In contrast, deaths and disease burden have decreased steadily in high income countries from 1990 to 2019.

Reproduced with permission © 2022, ERJ. See appendix: copyright permissions.

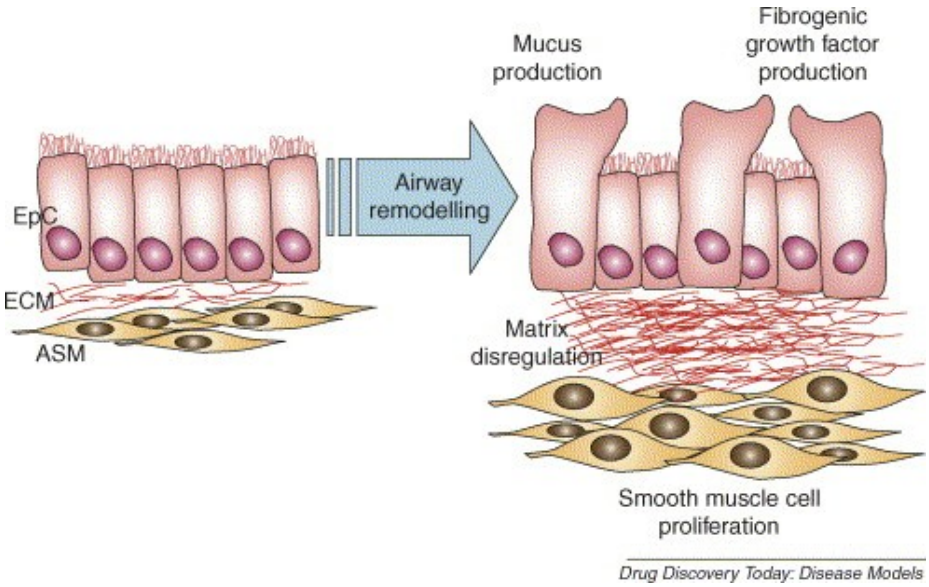
Estimated at 9166.57 per 100,000, the prevalence of asthma in The United Kingdom is the second highest in the world (25). A health survey conducted in 2018 showed that close to 20% of the adult population in England have been diagnosed with asthma (25). In addition to the high prevalence of the disease, emergency admission and death rates are among the highest in Europe for childhood asthma (26). Moreover, asthma health care disparities continue to exist on the level of socioeconomic status and geographic regions in England (27). The high burden of

asthma in England, is most likely fuelled by unmet needs for effective asthma management, particularly socioeconomic deprivation with asthma care (27,28). This has spurred national efforts towards the improvement of asthma outcomes, with the National Health Service Long Term Plan (29) including respiratory disease (such as asthma) as a national clinical priority with the aim of early detection and diagnosis, improved outcomes, appropriate medication use and reduced exacerbations and emergency admissions.

1.3 The pathophysiology of asthma

The immunopathophysiology of asthma is characterised by chronic airway inflammation driven by both innate and adaptive immune responses (30). Several triggers of chronic airway inflammation have been identified and established as prominent risk factors of asthma. These include allergens, infections, smoking, cold air as well as hormonal changes, obesity, genetic mutations and exercise. The resulting airway obstruction and hyperresponsiveness are hallmarks of asthma, which involves swelling of the airways, mucus hypersecretion, mucus plugs and airway remodelling (**Figure 4)**.

Figure 4. Airway remodelling in Asthma (31)



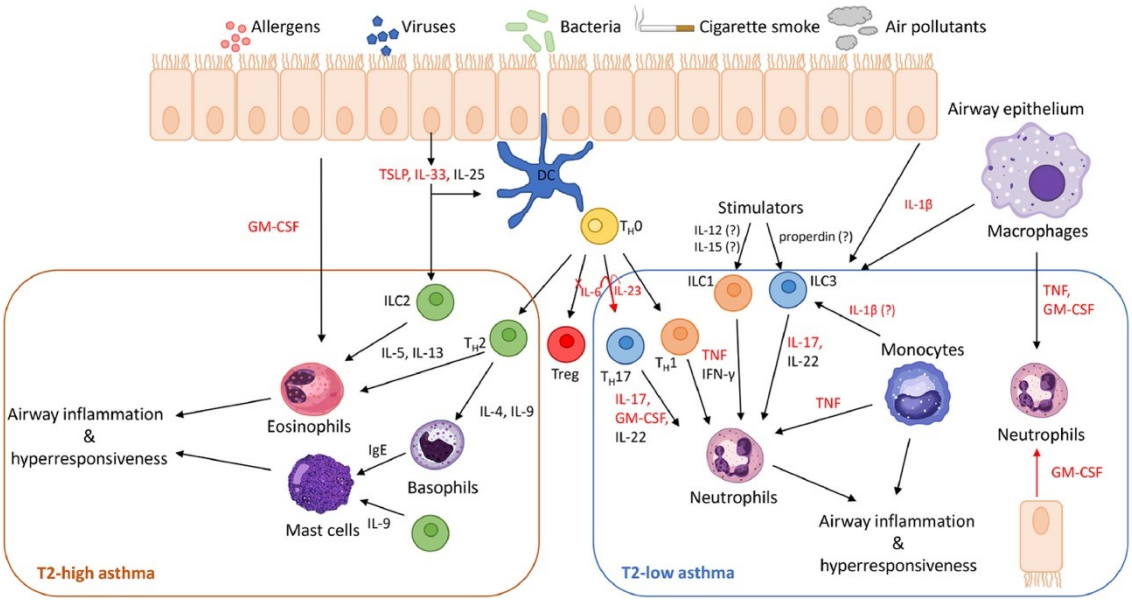
Legend: Chronic asthma is associated with structural changes to the airway wall which are termed airway remodelling. These structural changes include hyperplasia of the bronchial epithelial cells (EpC), which leads to excess mucus secretion and might lead to occlusion of

the airways. An imbalance in metabolism of extracellular matrix (ECM), either increased production or reduced breakdown of matrix proteins, leads to deposition of collagens and other matrix protein within the subepithelial and submucosal areas. Myocyte muscle mass is increased by hyperplasia airway smooth muscle (ASM) and myofibroblast proliferation. The combination of these processes leads to airway narrowing and therefore, reduced lung function.

Reproduced with permission © 2004, Elsevier Ltd. See appendix: copyright permissions.

As the name suggests, airway remodelling refers to the ongoing and possibly irreversible structural changes that occur in the airway of asthmatics; several processes modify the architecture of both small and large airways in such a way as to narrow the airway lumen and increase airway wall thickness (32).

Figure 5. Pathophysiology of T2-high and T2-low asthma (33)



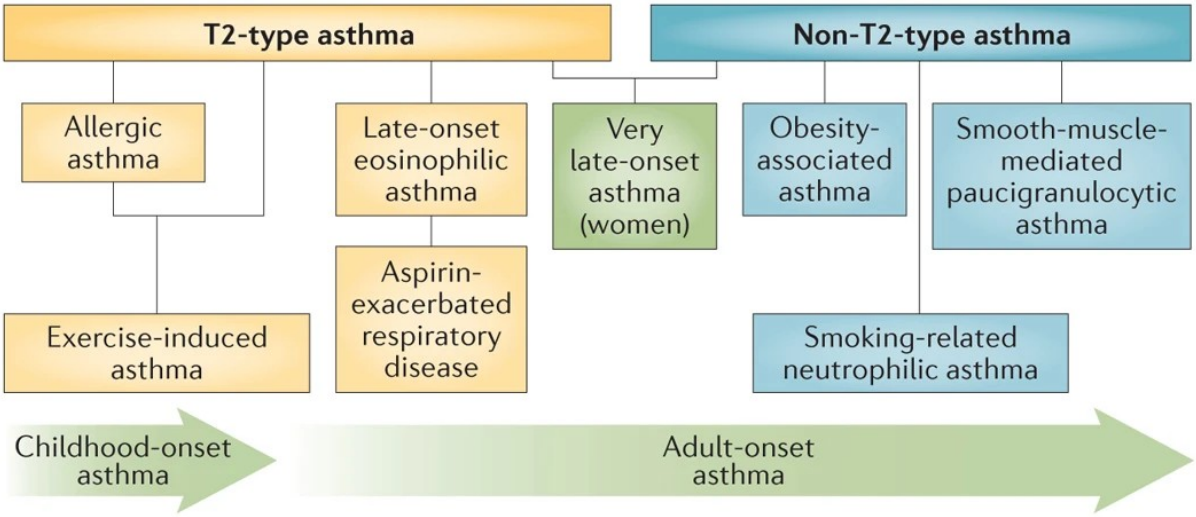
Legend: Alarmins secreted by airway epithelial cells activate T cells and innate lymphoid cells (ILC). Type 2 cytokines mediate the recruitment of eosinophils/basophils in T2-high asthma. Type 1 and type 17 cytokines are thought to drive T2-low asthma and contribute to neutrophil recruitment. Cytokines discussed in this review are indicated in red. DC, dendritic cell; GM-CSF, granulocyte-macrophage colony-stimulating factor; IgE, immunoglobulin E; IL, interleukin; Th, T helper cell; TNF, tumor necrosis factor; TSLP, thymic stromal lymphopoietin.

Reproduced with permission © 2022, Elsevier Ltd.. See appendix: copyright permissions.

Airway inflammation is a major feature of asthma and is driven by a combination of T helper (Th) 1, Th2 and Th17-mediated responses (**Figure 5**) as well as genetic predisposition. While the Th1 response is usually involved in immune reactions to viral infections, it can also play a role in airway inflammation. That being said, Th2-mediated immune reactions, better known as type 2 (T2) inflammation, are observed

in the majority of paediatric and adult asthma (**Figure 6**) cases involving sensitization to allergens (34–36). Once activated by dendritic cells in reaction to allergens in the airway (**Figure 5**), Th2 cells will initiate a cascade leading to bronchoconstriction starting with the release of Th2 cytokines such as interleukin (IL)-4, IL-5, IL-9, and IL-13. Epithelial cell-derived cytokines or alarmins such as IL-25, IL-33, and thymic stromal lymphopoietin (TSLP) perpetuate the inflammatory mechanisms underlying eosinophilic allergic asthma (37,38). Clinically, readily accessible biomarkers such as blood eosinophil counts and fractional exhaled nitric oxide (FeNO) are used to define T2 asthma with levels of ≥ 150/µl or ≥ 25 ppb respectively according to the GINA Guidelines (36).

Figure 6. Clinical phenotypes of asthma as T2-high and T2-low endotypes (39)



Legend: Depiction of the T2 and non-T2 associations with the clinical phenotypes of asthma.

Reproduced with permission © 2015, Springer Nature. See appendix: copyright permissions.

While this immune response is mainly mediated by eosinophils, other immune cells can be involved, namely neutrophils, macrophages, mast cells, among others (40) (**Figure 5**). T2 inflammation is predominantly found in classic allergic-type asthma which involves mast cell activation (13), but neutrophilic asthma has also been described and was linked with Th1 (41), Th17 cells (42) and with IL-6 trans-signalling (43). In addition to its prominent positive drivers of inflammation, asthma involves defective termination or resolution of inflammation culminating in failure to downregulate inflammatory responses. This leads to prolonged survival of

inflammatory cells such as mast cells and eosinophils in the airway, and by extension, continued airway inflammation. Within T2-low inflammation IL-17 has been recognized as a potential contributor to neutrophilic asthma (44) with an IL-17-driven non-T2 molecular endotype of severe asthma reported in 2019 (45). However associations with the Th17 pathway remain unclear with an analysis of blood, induced sputum and airway tissue showing no clear link of Th17 cells with asthma in general or specifically with neutrophilic asthma (46,47). This suggests that other factors may play an important role in asthma pathophysiology. The following section will explore the airway microbiome and its potential implications on asthma with a glance on advancements in other respiratory diseases.

1.4 The Airway Microbiome

1.4.1 The airway microbiome

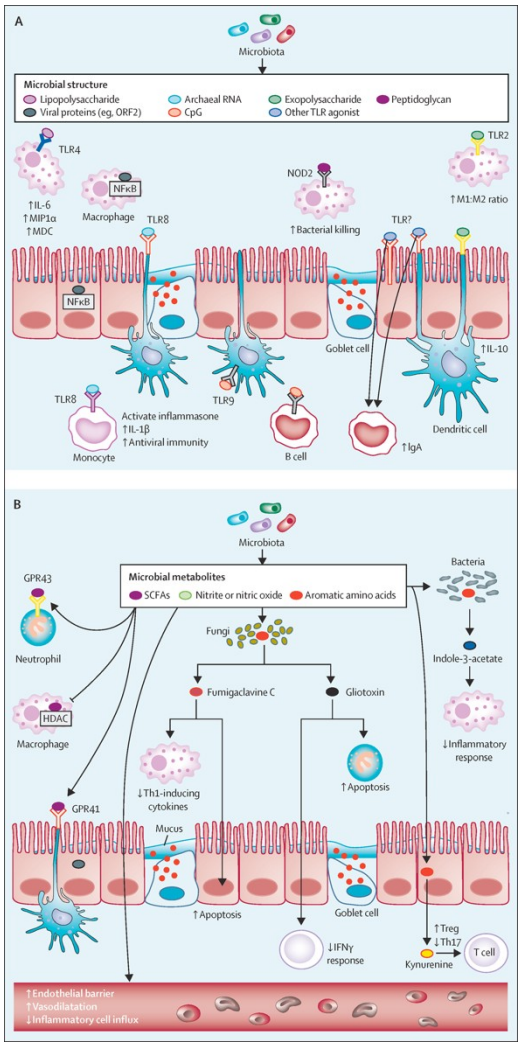
The human lung, which was previously thought to be a sterile environment below the larynx (48), is now known to harbour a variety of different types of micro-organisms including *bacteria*, *archaea*, *fungi*, *protozoa* and *viruses* (49,50). These microbes and their interaction with their local environment are thought to play an important role in respiratory health and disease. The symbiotic co-evolution of humans and microbes has fostered an ecosystem which is finely tuned to provide balance with fluctuations in host physiology contributing to the maintenance of homeostasis (51). The respiratory microbiome experiences variations in its composition in response to different factors, such as breastfeeding, antibiotics, living and environmental situations (e.g. house dust mites, allergens, second-hand smoke exposure), vaccinations and infections (52). The composition of the microbiome also varies by specific location in the lung, with lower levels of the respiratory tract associated with decreasing microbial density (53). In normal “healthy” circumstances, the genera *Streptococci* will dominate the upper airway microbiome, followed by *Neisseria*, *Prevotella*, *Rothia*, and *Haemophilus* (54). Similarly, variations in the upper airway virome and mycome are evident albeit with the general abundance of the *Herpesviridae* family in the virome (55), and *Candida* and *Saccharomycetales* in the mycome. As for the lower respiratory tract, the microbiome is generally transient as

introduced microbes are cleared by innate systems. *Bacteroidetes*, *Proteobacteria*, *Firmicutes* compose the bacterial microbiome of the lower respiratory tract, and at genus level, *Streptococcus*, *Prevotella* and *Veillonella* are described (56). The virome is mainly composed of *Anelloviridae* while *Ceriporia lacerate*, *Saccharomyces cerevisiae* and *Penicillium brevicompactum* make up the mycobiome (57,58).

The human gut is the most densely colonised organ, inhabited by 100 thousand to 100 billion bacteria per mL of luminal content (59). Gut bacteria serve several functions that are beneficial to the host, such as production of vitamins, ion absorption, immune function enhancement and even protection against pathogens (60). Both human and animal studies have provided confirmation of the connection between the lungs and the gut (58). The gut-lung axis was further supported by the evident influence between different lung diseases and the intestinal microenvironment (61). The opposite relationship also holds true. In asthma, the interactions between the gut and the lung are largely attributed to their microbiota (62). Similar associations and changes in gut microbiota were reported from other respiratory diseases, such as severe bacterial pneumonia, cystic fibrosis and influenza when compared to healthy individuals (62–64).

Dysbiosis of the gut microbiota in early life has also been implicated in the development of many respiratory diseases through the modulation of immune cell maturation and resistance to pathogens (65). Children who were shown to have significantly reduced relative abundance of some bacterial taxa in the first 3 months of age were at higher risk of asthma compared to their healthy counterparts (66–68). This may be due to the hypothesis that a diverse microbiome promotes immunological tolerance, thus with the reduction of microbiome diversity early in life, the immune system moves toward a state that promotes allergic disease and asthma (66). As compared with the gut microbiome, the airway microbiome is more sparse, exhibits a low biomass and is more amenable to shifts in composition (50). However, the interaction of airway microbiota with host immunity and inflammatory responses in the lung suggests that similar to the gut, the lung microbiome plays a role in modulating clinical and inflammatory asthma phenotypes (**Figure 7**) (67,68). This will be discussed in more details in the next section.

Figure 7. Signals from the respiratory microbiome influence lung immunity (67)



Legend: Respiratory microbiota produce structural ligands (A) or metabolites (B), which influence host immune activity in health and disease (arrows). GPR=G protein-coupled receptor. HDAC=histone deacetylase. TMA=trimethylamine. SCFA=short chain fatty acids. TLR=toll-like receptor. MDC=macrophage-derived chemokine.

Reproduced with permission © 2019, Elsevier. See appendix: copyright permissions.

1.4.2 Lung microbiome in respiratory diseases: Focus on asthma

Microbes are known to play an important role in all ecosystems they engage in. Among asthmatics, studies have shown the bacteria *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis* to be clinically relevant contributors to asthma exacerbations (69,70), although these species have also been cultured in stable in asthmatics (71). Compared with a healthy subject, the

emergence of these bacteria suggests a dysbiosis of the lung and potentially gut microbiome (72). This dysbiosis may be implicated in the development and inception of asthma (63,73). The hygiene hypothesis proposes that the environment one is exposed to in early life may influence the development of the microbiota in the lungs (74). Given their proximity to the mucosal membrane, these species may play a role in the shaping of the immune system; less exposure to pathogenic bacteria during infancy may result in the development of allergies later in life.

The lung microbiota have actually been linked to both childhood and adult asthma; neonates with detected presence of pathogens such as *S. pneumoniae, M. catarrhalis* and *H. influenza*e, had a higher risk of childhood wheeze and asthma at the age of 5 years (75). Recurrent wheeze and persistent childhood asthma has been confirmed by 16 rRNA gene sequencing to be linked to the dominance of *Staphylococcus* in the upper respiratory tract during early childhood (76,77). The role of the lung microbiome in the development of childhood asthma is further supported by the sharp increase in the risk of asthma among children who received one or more antibiotic courses in early life (78–80), most likely due to the reduction of species diversity caused by antibiotics (81,82).

Similar associations have been reported in adult asthma. Asthmatics have a higher bacterial burden compared to healthy controls (83,84). While keeping in mind that sampling site (i.e. between bronchial brushing, induced sputum, oral wash, nasal brushings) can affect results, studies have reported that mild asthmatics who are not being treated by ICS have reduced abundance of commensal organisms (*Prevotella* and *Veillonella*), but increased abundance of *Proteobacteria*, mainly *Haemophilus*, *Neisseria* and *Moraxella* (56,83–85).

In addition to its association with childhood-and adult-onset asthma, the airway microbiome can provide important insights into both asthma phenotypes and endotypes. In regard to asthma phenotypes, the *Enterobacteriaceae* and *Actinomycetaceae* families are common in asthmatics and are associated with the eosinophilic inflammatory phenotype (86). The examination of microbiota in severe asthmatic patients' sputum allowed Abdel-Aziz *et al* (87) to identify 2 microbial clusters based on Bray-Curtis β-diversity measures. Patients belonging to the cluster with decreased microbial richness and diversity had associations to worse lung

function, blood or sputum neutrophilia as well as higher likelihood of requiring add-on asthma treatment compared to those in the other cluster. The cluster with higher microbial richness and diversity was also detected in mild to moderate asthma, confirming the correlation between lung microbiota and asthma severity (87). Increased relative abundance in different microbes was linked to either having allergic asthma (atopic asthma) or asthma without atopy; Increased relative abundance of *Prevotella*, *Actinomyces* and *Lactobacillus* was associated with allergic asthma while asthmatics without atopy had an increased abundance of the genera *Aggregatibacter*, *Haemophilus* and *Actinobacteria* (85).

As for asthma endotypes, decreased bacterial burden and diversity has also been associated with T2-high asthma, while T2-low samples had an increased relative abundance of *Haemophilus*, *Moraxella* and *Neisseria* (88). The AMAZES study by Gibson *et a*l suggested that macrolide antibiotics may be preferable for use before OCS in severe T2-low asthma (89). As previously mentioned, eosinophilia in asthma has been linked to increased abundance of certain microbes, namely the *Halomonas*, *Aeribacillus*, *Streptococcus* and *Actinomycetaceae* genera (70,90). On the other hand, the severe treatment resistant neutrophilic asthma endotype has been linked to abundance of the genera *Moraxella*, *Haemophilus* and *Streptococcus* in the sputum (91). Based on extensive and increasing evidence of the lung microbiome's involvement in asthma pathogenesis and clinical presentation, it stands to reason that the microbiome could constitute a valuable and potential target for asthma prevention and treatment (92), as will be discussed in the next section.

1.5 Asthma Treatment: current approaches and the respiratory microbiome

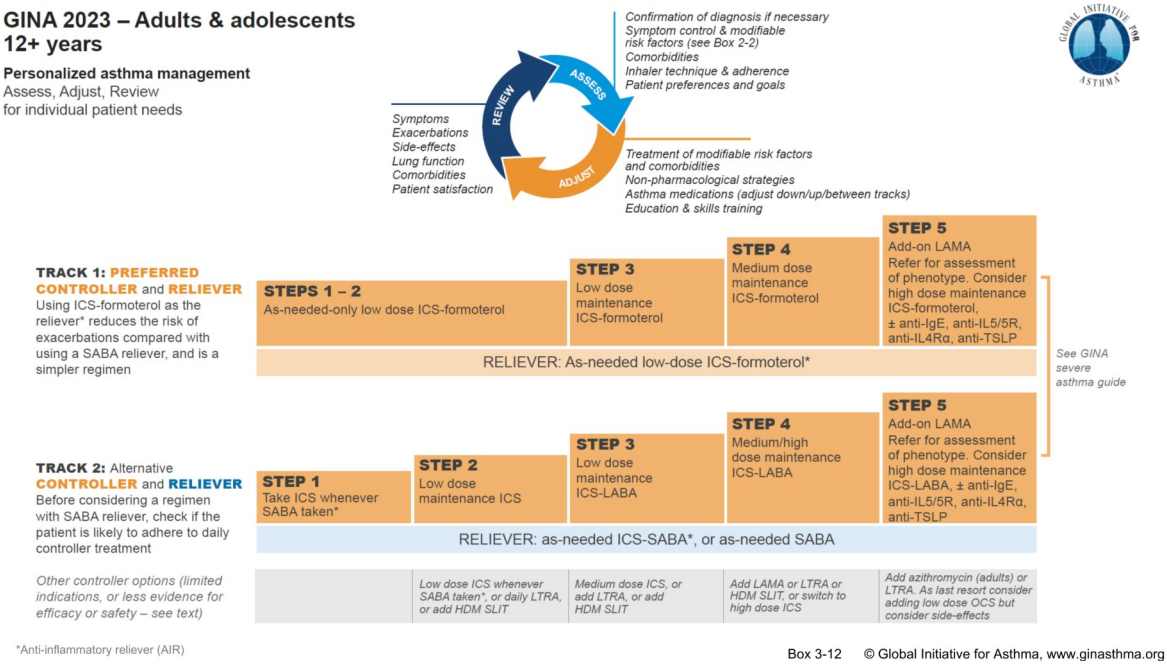
1.5.1 Conventional asthma treatment

The latest GINA guidelines, published in 2023 (93), maintain that the main goals of asthma treatment are symptom control and prevention of future asthma-related complications (mortality, exacerbations, airflow limitation, and treatment side effects). Asthma treatment follows a continual cycle of assessment, adjustment of treatment and review; assessment focuses on asthma symptoms, exacerbations, lung function and treatment side effects, in addition to patient adherence, inhaler technique,

comorbidities and preferences (94). Treatment is then adjusted based on these assessments, be it based on pharmacological or non-pharmacological strategies. Following this, a review of response to treatment is done in keeping with treatment goals. The GINA guidelines encourages shared decision-making when it comes to asthma management, considering the crucial role of patient involvement in improving treatment outcomes (95).

Several types of medications are used for asthma treatment (93). ICS are the hallmark of first-line asthma management. Daily use of ICS was shown to act on asthma's underlying inflammation and reduce cell infiltration of the lung epithelium, swelling and mucus production (96). Corticosteroids are effective against allergic and T2-driven asthma exacerbations (96). A leukotriene receptor antagonist (LTRA) can also be used, although it is less effective than ICS (97). To note, a LTRA may be particularly useful in Th2 and Th17-driven asthma, but not Th1-driven asthma. LABA are another major class of medication used for asthma treatment. They are usually prescribed in combination with ICS based on their action on smooth muscles in the airway and their relief of bronchoconstriction. Another add-on treatment option to ICS, LABA and LTRA is long-acting muscarinic agents (LAMAs). LAMA also reduces bronchoconstriction through its antagonistic activity on acetylcholine, with a possible role in neutrophilic asthma.

Figure 8. Global Initiative for Asthma treatment guidelines (93)



Legend: The Global Initiative for Asthma (GINA) guidelines for recommended steps in therapy in adults and adolescents aged 12+ years.

Reproduced with permission © **2021, Global Initiative for Asthma, available from ginasthma.org, published in Fontana, WI, USA. See appendix: copyright permissions.**

There are two basic treatment "Tracks" (**Figure 8**) recommended for adult and adolescent asthmatics, both of which involve stepping-up treatment as needed (1). Simply put, the preferred treatment approach (referred to as Track 1), relies on low-dose ICS-formoterol as a "reliever" medication. In other words, an ICS-containing treatment is used whenever relief of symptoms is needed ('as-needed' therapy) in mild cases. Patients with more frequent or more severe asthma symptoms be prescribed regular daily ICS-formoterol as prophylactic therapy (93). The use of an ICS-containing treatment for asthma is preferred seeing as it reduces the risk of exacerbations and leads to comparable control of asthma symptoms compared to a short-acting beta-agonist (SABA) reliever alone (98). This clinical benefit is observed irrespective of the baseline severity, lung function or inflammatory profile of the patient (99,100). Moreover, the use of the same inhaler as a reliever and (if needed) prophylaxis is a simple and convenient approach for both patients and clinicians.

Track 2 is an alternative to Track 1 wherein SABA alone or in combination with ICS is used as reliever medication (93). However, in addition to the higher risk of

exacerbations and need for emergency care associated with a SABA-only approach, it also promotes the risk of poor adherence to ICS once prescribed (93).

Both Track 1 and Track 2 recommend the consideration of biologic therapy at severe (Step 5) asthma. Patients at this stage should be referred to a specialist for assessment, phenotyping and subsequent biologic therapy tailored to each patient's asthma phenotype (93). Anti-immunoglobulin E treatment should be considered for patients with allergic asthma (101), while those with eosinophilic asthma would benefit from anti-IL5/5R treatment (101–103). In a similar vein, anti-IL4Ra therapies are recommended for severe eosinophilic/ T2 asthma cases as determined by high blood eosinophil (≥150/μl) or FeNO ≥25 ppb levels or OCS-dependent patients (101,104–106), and anti-TSLP treatment can be considered for the management of severe asthma in patients aged 12 years or above (101).

Severe asthma outcomes remain suboptimal despite the use of biologic therapies, creating a need for new techniques such as precision medicine.

1.5.2 The lung microbiome and asthma treatment

While there is limited to no data on bronchodilators' effect on lung microbiota, there is some evidence to suggest that the LABA formoterol and the LABA/ICS combination treatment can cause a decrease in microbiome diversity in patients with chronic obstruction pulmonary disease (COPD), possibly affecting response to bronchodilators (107). Evidence on corticosteroids is more prevalent, with studies showing changes in microbial diversity and composition based on the use of ICS (108,109); an increase in the abundance of *Proteobacteria* has been linked to ICS use (80), while ICS-naive asthmatics have enriched *Haemophilus*, *Neisseria*, *Fusobacterium* and *Porphyromonas* (110). The airway microbiota can also change in response to ICS dose, with *Streptococcus* enrichment associated with lower doses, whilst *Parvimonas* and *H. parainfluenzae* were linked to higher ICS doses (110,111). Similar influences were noted between oral corticosteroid use and the lung microbiome (110).

Biological therapies targeting T2 inflammatory pathways can also affect the lung microbiota (112). The use of omalizumab was linked to a decrease in *H. influenza*

and *M. pneumonia* and *Bordetella*-free bacteria pellets but an increase in *C. pneumonia*, *N. meningitidis* and *S. aureus* bacteria pellets (112). A significant modification of the airways microbiome is reported with biologics, as in the case of mepolizumab (113). In addition, an alteration of the lung microbiota has been observed two years post treatment with reslizumab, particularly an increased abundance in *Mycoplasma pneumonia*, *Neisseria meningitidis* and *Haemophilus influenza* (112). Evidence from the sputum of COPD patients showed that the biologic benralizumab can reduce bacterial load, with *Streptococcus pneumoniae* being significantly reduced compared to placebo, however no changes in clinical outcomes were observed (114).

Further exploration of the lung microbiome remains sorely needed in asthma, particularly how it may be utilised as a therapeutic target for the improvement of clinical outcomes.

1.6 Phenotypes and endotypes: A place for molecular phenotyping in asthma

1.6.1 Asthma phenotypes:

Understanding and recognizing the biological and immunological processes in asthma is crucial for the characterization of asthmatics and guiding clinical management (19). Traditionally, asthma is divided into phenotypes based on demographic, clinical and pathophysiological characteristics (19). These phenotypes ultimately fail to account for disease mechanisms and consequently have limited practical utility. Asthma phenotypes recognized by GINA are based on clinical features and represent allergic asthma, non-allergic asthma, adult-onset or late onset asthma, asthma with persistent airflow limitation and asthma with obesity (93). These phenotypes are not necessarily distinct and may overlap. Early-onset (childhood) asthma is generally attributed to allergic asthma, which is also eosinophilic and responsive to ICS (115). A more heterogeneous presentation is evident with non-allergic asthma, which could involve eosinophilic, neutrophilic, pauci-granulocytic or even mixed inflammatory subtypes, with worse response to ICS-based therapy (114). While males are more likely to present with early-onset allergic asthma, adult-onset asthma occurs more frequently in females and also has

a poor response to ICS as seen in non-allergic asthma (114). Obesity-related asthma is another recognized phenotype commonly associated with persistent airflow limitation and airway remodelling fuelled by neutrophilic, rather than eosinophilic inflammation (115,116).

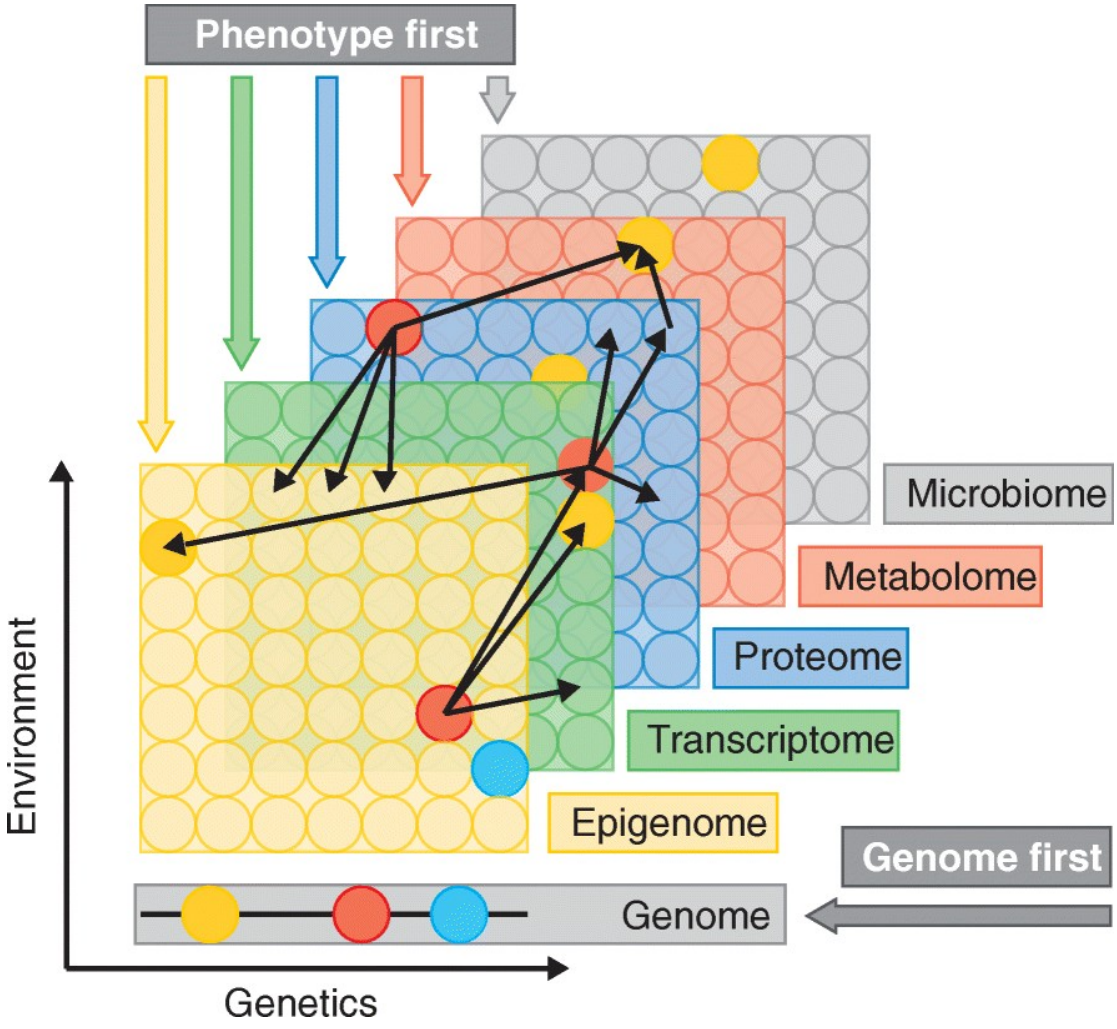
1.6.2 Asthma endotypes

Considering the limitations of asthma clinical phenotypes in delineating asthma mechanisms, asthma molecular phenotypes or "endotypes" are now recognized in order to optimise disease management and further support the use of personalised and precision medicine (117). Asthma endotypes would divide asthma into distinct disease entities with specific mechanisms that can be targeted with precision medicine (12,118–120). The identification of the T2 inflammatory complex endotype was the most notable advancement in this regard with the enhanced mRNA expression of periostin, serpin B2 and CACL1 in bronchial epithelial cells following IL-13 treatment and in a subset of patients with asthma (121,122). These patients responded well to ICS therapy and eventually led to the expansion of treatment options for previously poorly treated/controlled patients. Biologics targeting the T2 inflammatory pathway have paved the way for better outcomes in patients with severe asthma with a high number of exacerbations, high blood eosinophil counts and/or a high FeNO level. These patients were often relatively refractory to corticosteroids and other standards of care, yet were clinically indiscernible from their corticosteroid-sensitive counterparts (122,123). The use of targeted therapy had another distinct advantage, namely their corticosteroid-sparing or -tapering effect in patients who are dependent on these therapies for symptom control yet continuously suffering from their side-effects (124). However, initial studies using mepolizumab and other anti-IL5 monoclonal antibodies (moAbs) failed clinically in asthmatics not selected for T2 high features (i.e. FeNO and blood eosinophils) and using primary endpoints such as FEV_1 rather than exacerbation rate (125), despite suppressing blood and sputum eosinophils.

1.6.3 Molecular phenotyping: advancements and gaps

The ‘omics’ suffix has emerged to describe a comprehensive assessment of a particular type of molecule (126). This started with genomics, which was coined in 1986 (127), but now incorporates other molecular types including transcriptomics, epigenomics, proteomics, metabolomics and microbiomics (126). Analysis using single-omics alone is limited as it can only provide correlations or associations within that single molecular group or with phenotypes; Integrating multi-omics datasets together can provide a greater understanding of the interaction between these individual omic datasets and insights into the mechanisms at play **(Figure 9)** (126). This integrative approach is especially important when it comes to disease with a complex aetiology, where a clear deterministic factor is missing (126,128).

Figure 9. Multiple omics data types and approaches to disease research (126)



Legend: Layers depict different types of omics data. Omics data are collected on the entire pool of molecules, represented as *circles*. Except for the genome, all data layers reflect both genetic regulation and environment, which may affect each individual molecule to a different extent. The *thin arrows* represent potential interactions or correlations detected between molecules in different layers—for example, the red transcript can be correlated to multiple proteins. Within layer interactions, although prevalent, are not depicted. *Thicker arrows* indicate different potential starting points or conceptual frameworks for consolidating multiple omics data to understand disease. The genome first approach implies that one starts from associated locus, while the phenotype first approach implies any other layer as the starting point. The environment first approach (not shown) examines environmental perturbations.

Reproduced under creative commons license © **2017,** Springer Nature. **See appendix: copyright permissions.**

Given the complex disease aetiology of asthma (128), Integrative omics can provide an important platform for improving the understanding of asthma endotypes and developing optimal treatment approaches (129). Previous applications of integrative multi-omics have succeeded in generating much-needed insights into the mechanisms and host-microbe interactions underlying another major respiratory condition, COPD (130,131). A similar approach is therefore needed in asthma, particularly for severe asthma where response to therapy remains suboptimal even with the use of biomarkers and biologics. Asthma management strategies would benefit from a deeper evaluation of disease mechanisms, particularly the host-microbiome interactions that could be driving inflammatory responses (15).

Multi-omics integrative approaches have been categorised into three groups by Hasin *et al* in 2017 (126); “genome first”, “phenotype first”, and “environment first”. For this thesis, “phenotype first” and “environment first” are most relevant. “phenotype first” approach looks at how different in phenotype of disease, such as disease severity, correspond to changes in the omics data and likewise “environment first” looks at how factors such as diet perturb different layers of omics data (126).

With the advent of high-throughput sequencing techniques such as next-generation sequencing (NGS) (132) and the continued price reduction of these techniques, metagenomic sequencing of microbial communities’ resident in human microbiomes has become possible.

Around 3 decades ago, culture-independent studies assessed the microbiota community using targeted sequencing of the highly conserved 16S ribosomal ribonucleic acid (rRNA) gene (133). In turn, this has allowed the development of more rapid methods of surveying the microbiota directly from deoxyribonucleic acid

(DNA) extracts (131). In asthma, the majority of analysis undertaken have utilised 16S small subunit ribosomal RNA (rRNA) - 16S rRNA based profiling of human microbiota (56,70,83,85,86,90,91,110,134–138).16S-based profiling identifies the 16S small subunit rRNA gene which provides a targeted bias approach while metagenomics can provide an unbiased approach to sequence the whole microbiome (139). However, for this to happen, metagenomic sequencing pipelines needs to be evaluated and selected to parse the sequence data and remove noise.

The evolution of these high-throughput methods has generated vast amounts of data (140). In 2012 Dennis Noble proposed that there is no privileged level (omics layer) of causation; thus biological systems cannot be determined by genomics alone but rather by the interactions between each layer (141) . This has driven the need for conceptual approaches that analyse these system-wide datasets in a holistic manner providing details on complex interactions and behaviours which emerge from systems properties of these samples (129). Unsupervised integration or fusion of these heterogeneous omics datasets to create a comprehensive view of the interactions which drives a particular disease system through the extraction of homogeneous latent factors, is becoming increasingly important (142). These methods aim to cluster patients together into meaningful disease subtypes which will form the basis of distinct tailored therapies (142). Several integrative and clustering methods exist which fall into three broad categories; matrix factorisation or regression-based analysis, set-based analysis and network-based analysis (143–149).

Matrix factorisation aims to infer a set of latent variables in a lower dimensional space from data by find factors of a data matrix consisting of different omics layers. It involves the factorization of a non-negative matrix of omics variables into two non-negative matrices, matrix and matrix :

If dimensions the factorisation is lossy and hence is formulated as such:

and is formed as an optimisation problem and achieved by minimising error:

This method has an inherit clustering property as it extracts a low number of common latent features from multiple high-dimensional data into latent space W and the columns of H give the contribution of each basis. The latent space represents the commonality and may account for distinct molecular phenotypes related disease features. The iCluster method (144) formulates the clustering as a joint latent variable that is estimated from multiple omic layers where the clustering scheme in each layer is represented as a Gaussian latent variable model.

Canonical Correlation Analysis (CCA) is a method to identify interactions between two datasets. CCA searches to find linear combinations of variables from the two datasets that maximizes their correlation with each other in order to find loading weights for each component (a pair of orthogonal canonical variates):

The number of components is equal to the minimum number of variables in either or . The dimensionality of the maximisation problem can be reduced by applying a lasso penalty.

The application of integrative omics in asthma is critical considering the limitations of current treatment approaches. However, it requires the understanding of the airway microbiome and its complex interactions before any asthma subtypes can be identified and targeted. The following section will explore the airway microbiome and its potential implications on asthma with a glance on advancements in other respiratory diseases.

1.7 Towards bridging clinical gaps: Precision medicine in asthma

Exploring the intrinsic biology of a disease and attempting to deepen the understanding of its mechanisms, pathogenesis falls in line with calls for the establishment of a new disease taxonomy and the improvement of healthcare outcomes (150). Despite the integration of biologics into the management of asthma,

significant clinical gaps are still faced as many patients have poor asthma control (151). Ongoing research is critically exploring clinically relevant biomarkers for the easy and low-cost determination of optimal patient candidates for each biologic, particularly for the as-of-yet enigmatic non-T2 asthma (152). It is important to note that the distinction between T2 and non-T2 asthma might ultimately be an oversimplification as it is not possible to fully reconcile the clinical response to biologics of many severe asthmatics with the carefully crafted asthma phenotypes (19). New therapeutic approaches are thus likely to depend on an individualised approach that consists of targeting “treatable traits'' and “treatable mechanisms” specific to asthma phenotypes and endotypes in each asthma patient (153). To achieve this, researchers and clinicians need to mobilise the power of an innovative approach to medical management tailoring: precision medicine (also referred to as personalised medicine, tailored medicine, stratified medicine, or targeted medicine) (153).

Rather than the traditional one-size-fits-all approach where treatments are developed based around an average response of a population, precision medicine harnesses the multi-omics data to tailor treatment decisions and medical interventions based on an individual’s lifestyle, genes and environmental exposure (150). Precision medicine aims to treat “the right patient, with the right drug, at the right time” (154).

For this to be achieved in the clinic, asthma needs to be understood as a complex dynamic system with many interacting parts (15). Harnessing multi-omics data and systems biology becomes critical for the analysis, modelling and understanding of these complex interactions. Defining asthma endotypes and treatable traits requires the analysis of all genes, proteins, lipids, metabolites and factors that might be implicated in the pathogenesis of asthma. Precision medicine is not limited to investigating interactions on the level of patients’ genome, but also considers the impact of the exposome (environmental exposures) and the microbiome (exposure to pathogens) on cellular and immune function (63). With the employment of systems biology’s integrative approach, asthma pathways and processes could be unveiled through the modelling of these intricate systems with multi level multi scale mathematical and computing methods (155). These methods will rely on data generated from multi-omics levels, such as transcriptomics or proteomics which

provide information regarding changes in the host immune/inflammatory response associated with alterations in the micriobiota.

As previously mentioned, current trends in asthma management favour the distinction of treatable traits based on clinical phenotypes and endotypes (156). Currently recognized treatable traits in asthma are divided into three domains, each with their distinct treatment approaches (157): the pulmonary domain encompasses airway eosinophilic inflammation, airway hyperresponsiveness, proclivity to lung attacks, mucus hypersecretion and pathogen colonisation particularly with *Haemophilus influenzae* (157); the extrapulmonary domain is characterised by osteopenia, sarcopenia, activity limitation and dysfunctional breathing (157); and the behavioural domain includes patients with suboptimal self-management skills (poor inhaler technique, inhaler use, planning for asthma attacks) (157) and those who are current smokers but willing to quit (157).

Another notable distinction made in recent asthma research is the concept of treatable mechanisms, which are established based on molecular phenotypes (15). While treatable traits describe clinical features based on clinical phenotyping, effective treatments are not always readily available for these “treatable” traits as inherently implied in their qualification (15). This gap can be bridged by identifying treatable mechanisms of the treatable traits; for example, the treatable trait of airflow obstruction could be due to differential expression of gene networks that are in turn associated with distinct inflammatory processes, airway remodelling and treatment responses (15).

The recognition of microbiological treatable traits in asthma is of particular note to this thesis. Prominent traits reported in asthma and COPD are chronic respiratory infection and microbiological colonisations (158). Integrative multi-omics have already been used to determine the mechanisms and host-microbe interactions in COPD (130,131). The presence of an organism in the human airway carries significant prognostic implications even in the absence of pathogenicity (159). A comprehensive evaluation of asthmatic patients’ microbiome composition is therefore crucial and is enabled by advanced culture-independent techniques, such as metagenomics (160).

By applying such techniques, it becomes possible to link changes in the airway microbiota to clinical asthma presentation. For example, an increase in the frequency of proteobacteria in the airway of asthma patients (56) has been associated with airway hyperresponsiveness and poorer asthma control (83). More specifically, patients with neutrophilic asthma were found to have reduced sputum bacterial diversity compared to other asthma phenotypes, as well as higher levels of *Proteobacteria*, including *Moraxella* and *Haemophilus* (particularly *Haemophilus influenzae*) (134,161). Studies of microbiome composition have also shown associations between stable asthma and concentrations of bacteria such as *Streptococcus pneumoniae*, *Moraxella catarrhalis* (75,91) and *Haemophilus influenzae* (56,83,91). These associations diversify treatment options for asthmatics, potentially ensuring much-needed improvement in asthma outcomes; antibiotics, mucolytics, roflumilast, education, and inhaled interferon-β treatment become possible therapeutic options (158,162,163).

The overall hypothesis of this thesis is that species-level sputum metagenomic data can provide insights into asthma phenotypes and when integrated with host sputum transcriptomics can reveal microbiome-host interactions and mechanisms at play between childhood-onset and adult-onset asthma.

1.8 Thesis Aims

The aim of this thesis is to understand whether species-level metagenomic data coupled with host-transcriptomics can better elucidate asthma patient phenotypes.

The objectives were:

- To characterize and investigate the microbiome of asthmatic patients down to the species level with the use of metagenomic whole genome shotgun sequencing approach from induced sputum.
- To better understand and explore the impact of relative dominant species (RDS) on the clinical phenotype of patients in addition to the effect on transcriptomic pathways.

- To investigate the interaction between the microbiome and the host gene expression in regard to the distinct mechanisms at play between childhood-onset and adult-onset asthma.

In the next chapter the datasets and methodologies used throughout the thesis are described.

Chapter 2. Datasets & Methodology

Parts of this chapter are based on the publication:

*Kermani N, **Versi A**, Gay A, Vlasma J, Jayalatha AKS, Koppelman GH, et al. Gene signatures in U-BIOPRED severe asthma for molecular phenotyping and precision medicine: time for clinical use. Expert Rev Respir Med. 2023;17(11):965–71.*

2.1 Introduction

This chapter introduces and describes the datasets used throughout the thesis in addition to the common methodologies employed throughout each of the chapters. Other methods that are specific for each Chapter are described in the appropriate chapters.

2.2 Datasets

2.2.1 U-BIOPRED

The Unbiased BIOmarkers in PREDiction of respiratory disease outcomes (U-BIOPRED) is a pan-European study conducted by a consortium of 20 academic institutions, 11 pharmaceutical companies and six patient organisations funded by the European Commission and The European Federation of Pharmaceutical Industries and Associations (EFPIA) (NCT01976767) (164). The study began in 2008 and was completed in 2013. The study was approved by the local ethics committee for each participating clinical institution, and adhered to the standards set by International Conference on Harmonisation and Good Clinical Practice. All participants gave written and signed informed consent.

It employs a bottom-up data-driven approach and was carried out in a non-interventional manner to discover biomarkers relating to asthma and to identify meaningful asthma subtypes.

2.2.2 Participants

The study produced a wide variety of multi-omics data for 610 clinically well-characterised adults, which were divided into four cohorts; severe asthmatic non-smoking subjects (SAn, n=311), severe asthmatic smokers and ex-smokers (SAs/ex, n=110), mild-moderate asthmatic non-smokers (MMA, n=88), healthy non-smoking controls (HC, n=101) (164)and 282 paediatric asthmatics (165). The analysis in this thesis is restricted to adult subjects. Details of the definitions used to define each group are provided in **Table 1**.

Table 1. Criteria for participants of each adult cohort in the U-BIOPRED study

Cohort A	SAn	severe non-smoking asthma	Participants in this group were non-smokers for at least the past 12 months, with a less than 5 pack-year smoking history, with asthma and uncontrolled symptoms defined according to Global Initiative for Asthma (GINA) guidelines (1) and/or frequent exacerbations (more than two per year) despite high-dose inhaled corticosteroids (1000 µg fluticasone propionate per day or equivalent dose).
Cohort B	SAs/ex	smokers and ex-smokers with severe asthma	This group was defined as for the SAn group except that they were either current smokers or ex-smokers with a smoking history of at least 5 pack-years.
Cohort C	MMA	mild/moderate non-smoking asthmatics	Participants in this group were non-smokers for at least the past 12 months, with a less than 5 pack-year smoking history and had controlled or partially controlled asthma symptoms, as defined by GINA, whilst receiving a dose of <500µg

			fluticasone propionate/day or equivalent.
Cohort D	HC	healthy non-smoking controls	These participants had no history of asthma or wheeze, had no other chronic respiratory disease, were non-smokers for at least the past 12 months with a smoking history of 5 pack years and their pre-bronchodilator forced expiratory volume in the first second (FEV1) was 80% pred.

2.2.3 Samples collected

The dataset consisted of samples taken from multiple compartments including nasal brushings, exhaled breath, sputum, bronchial brushings, bronchial biopsies, blood and urine. The type of data included metagenomic and microbiome 16S data (sputum) in addition to transcriptomic (blood, nasal and bronchial brushings, bronchial biopsies and sputum), proteomic (serum and sputum), lipidomic (serum and sputum) and metabolomic (sputum and urine) data.

Metagenomics and microbiome 16S data was collected from induced sputum samples and therefore the sputum compartment will be the focus of the thesis.

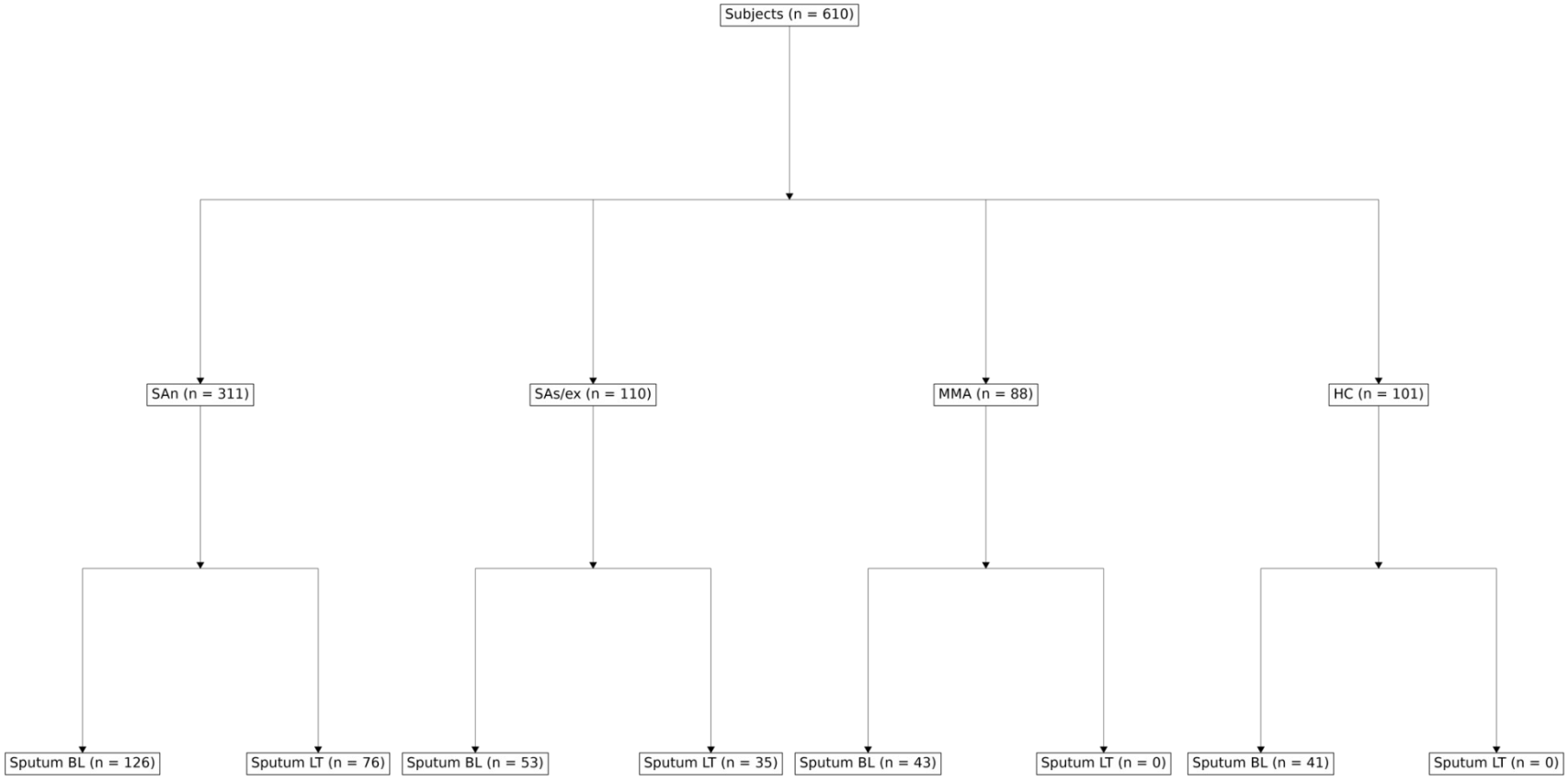
2.2.4 Sputum induction and initial sample handling

At baseline and at 12-16 months, induced sputum samples were obtained by inhalation of hypertonic saline solution, sputum plugs were collected and sputum supernatant and cells were obtained. The protocol for sputum induction and immune cell analysis has been published previously (166).

Induced sputum samples contain a composite of both the lower airway secretions in addition to pulmonary surfactant, saliva, immune and squamous cells from the upper airways. **Figure 10** shows a consort diagram indicating the number of samples from each subject group for which sputum samples were taken at each time point.

Longitudinal follow-up was performed only in patients with severe asthma and not in healthy subjects or patients with asthma classified as MMA.

Figure 10. Consort Diagram of Adult Cohorts with sputum samples collected and the timepoints of collection in U-BIOPRED



Legend: The consort diagram depicts the number adult participants in each cohort of U-BIOPRED and the number of sputum samples collected for both baseline and follow up.

Gene expression profiling was performed using Affymetrix microarray (GPL570 Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array) (Affymetrix, Santa Clara, CA) (167) with RNA extracted from sputum cells (GEO accession no.: GSE76262). This microarray assay chip contains over 47,000 gene specific probes, which when a probe is exposed to its corresponding sequence of mRNA, hybridisation of the mRNA to the probe occurs and quantification of the level of gene expression is provided.

Details of the RNA extraction, quality control and gene expression analysis have been reported elsewhere (168,169).

For metagenomic sequencing and 16S microbiome profiling, sputum samples were frozen and sent to Second Genome (San Francisco, California, USA). The frozen samples were treated with MoBio Tissue and Cells DNA Isolation Kit (Qiagen, Germantown, USA) according to the manufacturer's instructions to extract and purify genomic DNA. The DNA extraction was performed in a BioSafety Cabinet in an ISO 8 Cleanroom and each batch included negative and positive controls. Extracted gDNA was immediately stored in a -20C freezer.

For metagenomics sequencing, the frozen gDNA samples were prepared for sequencing with the Illumina Nextera kit (Illumina, San Diego, USA). The Nextera kit employs the technique of tagmentation, where unfragmented DNA undergoes fragmentation and adaptor tagging in the same step by utilising the transposase tn5 enzyme. The resultant tagged fragments were then amplified using polymerase chain reaction (PCR). Quantification was then performed with Quant-iT dsDNA High Sensitivity assays (ThermoFisher Scientific, Waltham, USA). Libraries were pooled and run using 100 base-pair paired-end sequencing protocols on the Illumina HiSeq® 2500 platform (170).

For 16S microbiome profiling, enrichment of the 16S v4 rDNA region was performed by DNA amplification using fusion primers designed against the surrounding conserved regions which are tailed with sequences to incorporate Illumina (San Diego, CA) adapters and indexing barcodes. Samples were then PCR amplified and those that met a quantification minimum were selected for pooling and sequencing. These selected samples were loaded into a MiSeq® reagent cartridge, and then onto

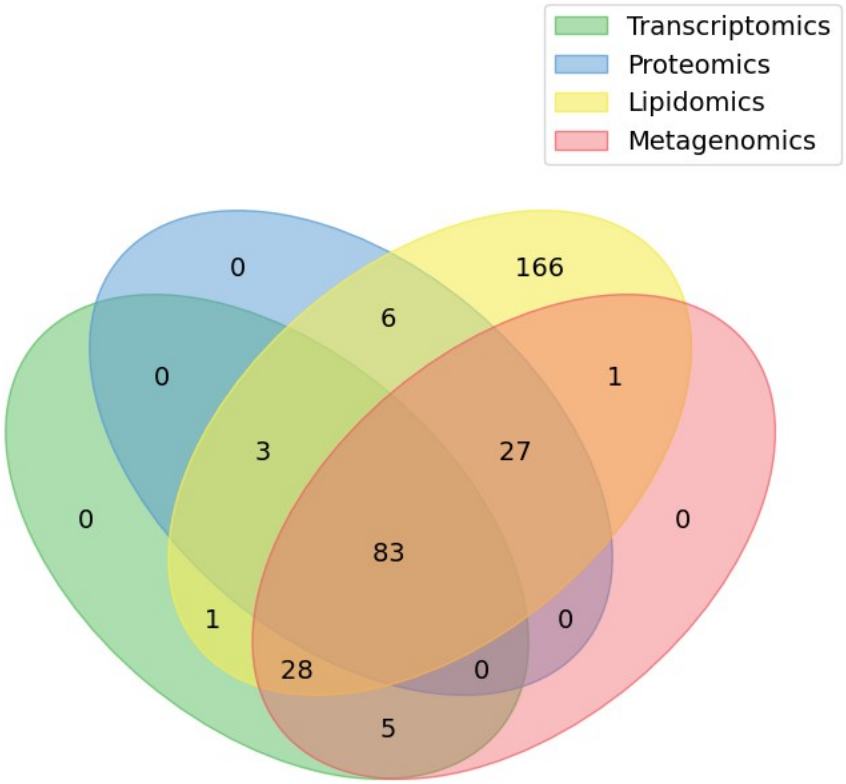
the instrument along with the flow cell. DNA fragments are amplified by a process known as cluster generation. After cluster formation on the MiSeq® instrument, the amplicons were sequenced for 250 cycles with custom primers designed for paired-end sequencing (171).

For lipidomics, lipids were extracted as described previously (172). Lipid extraction was performed using the semi-automated Bligh-Dyler protocol (173) and samples were analysed using shotgun mass-spectrometry (172).

Proteomics data was generated using the SOMAscan® V3 (174) proteomics assay (SomaLogic, Boulder, CO, USA; www.somalogic.com) on the sputum supernatant samples of the adult U-BIOPRED cohort as described previously (175). This technology is a high throughput, multiplexed proteomic assay, which simultaneously measures native folded proteins directly from small volumes of biological samples. It leverages the use of aptamers, which are single stranded oligonucleotides that are designed to fold into a designated shape in order to bind to specific target proteins (176). After the probes have been exposed to the sample, the number of each of these aptamer probes attached to its corresponding specific protein represents the abundance of that protein. These are then quantified as relative fluorescence units (177).

One of the limitations of the data is the overlap in the number of patients providing each of the multi-omics datasets. The Venn diagram **(Figure 11)** shows the number of patients for the baseline time-point which data has been collected for and is available for analysis after quality control (QC) and after bioinformatics pipeline have been run for each respective omics type. Collection for each patient for each omic dataset at both time-points would be ideal to allow for analysis between all the omic datasets. However rarely is this met due to the practicalities of taking multiple samples from patients and the processing of all these samples to identifying outliers during data pre-processing.

Figure 11. Venn Diagram depicting data overlap of available platforms in the sputum



Legend: The Venn diagram depicts the overlap of samples available at the baseline timepoint after sample collection, quantification and computational processing for the sputum compartment for transcriptomic, proteomic, lipidomic and metagenomic data platforms.

After quality checking and normalisation, datasets were uploaded on to the tranSMART system at the Data Science Institute (DSI), Imperial College London. TranSMART is an open-source knowledge management platform for sharing research data (178).

2.2.5 Statistical Methods

Data and statistical analysis in addition to pre-processing was performed using R version 4.1.1 (codename: Kick Things) (179) and pandas 1.4.2 (180) using python

3.9.10 (181). All code was run via Jupyter Lab, in an environment which was created using Docker (available here). Dependencies were pinned to ensure all analysis was performed under the same package versions. All notebooks were version controlled using Git. Intermediate datasets were version controlled using Git LFS and the processed U-BIOPRED data was accessed from tranSMART.

2.2.6 Differentially expressed gene (DEG) analysis

To identify the subset of genes which are differentially expressed between two groups of samples, a method known as differentially expressed gene (DEG) analysis was performed. This method compares gene expression levels between different groups or conditions to identify these differentially expressed genes. In general, the approaches used to identify these signatures typically start by defining a null hypothesis, generating a p-value, and then applying a significance threshold that considers multiple comparisons to minimize the risk of type 1 errors (182).

The most common techniques include Wilcoxon rank-sum test, Limma (183), edgeR (184) or DESeq2 (185). The three techniques of Limma, edgeR and DESeq2 are similar but have different steps and approaches. Limma is usually used for microarray continuous data and edgeR and DESeq2 for RNA-seq count data. Both Limma and edgeR start with normalised count data, often quantile normalisation or TMM (trimmed mean of M values) to control for sequence depth and RNA composition bias (184,186). Whereas DESeq2 uses an internal normalisation (size factor) to correct data for library size and RNA composition bias (185).

All three methods utilize empirical Bayes, calculating different methods of dispersion to moderate gene-specific variance to give more accurate identification of DEGs especially for genes with low or noisy counts. Gene expression values are then fitted to a model – Limma employs a linear model (for continuous data, although RNA-seq count data can be transformed via Limma-Voom) and t-statistic is used to find DEGs. Both edgeR and DEseq2 utilise a negative binomial generalised linear model (GLM) together with the Wald-test to find DEGs. These DEGs can then be used as a candidate gene signature.

In the U-BIOPRED sputum compartment, the focus of this thesis, Limma is utilized for DEGs as the transcriptomic dataset was derived from microarray techniques which generate continuous data as opposed to RNA-seq which generates count data. Unless stated otherwise, a $\log_2$ fold change of > 0.5 and an adjusted P-value of < 0.05 was used as a cutoff.

2.2.7 Pathway Enrichment

Pathway analysis methods for utilising gene signatures and assessing their enrichment in other samples include gene set enrichment analysis (GSEA) (187) and gene set variation analysis (GSVA) (188).

GSEA tests the hypothesis that none of the gene signatures is associated with the phenotype group. The genes are usually ranked by the mean expression level for the phenotype group and control group although other metrics can be used (189). Each gene in the gene signature is then matched up to the rank of the gene from the previous step. Then a walking sum is performed down the ranked gene expression data and a running sum of expression score is kept. If a gene is in the gene signature the aggregate gene expression value for that group is added to the running sum and if it isn't in the gene signature, then a set value is subtracted from the running sum. The peak value of the running sum is set as the enrichment score of that phenotype.

GSVA is used to calculate sample-wise enrichment (gene set score) of a set of genes (gene signatures) in normalised gene expression matrix. First the genes are ordered by rank of the expression levels for each sample. Then for each gene signature, the cumulative distribution function (CDF) of the ranked expression values of genes within the gene signature and genes not in the gene signature are calculated for each sample by means of a walking sum for both the within signature genes and the outside of signature genes. Kernel density estimation is performed to smoothen the CDF curves and finally the Kolmogorov-Smirnov test between the within gene signature CDF and outside gene signature CDF is calculated and the p-value gives the probability they come from the same distribution.

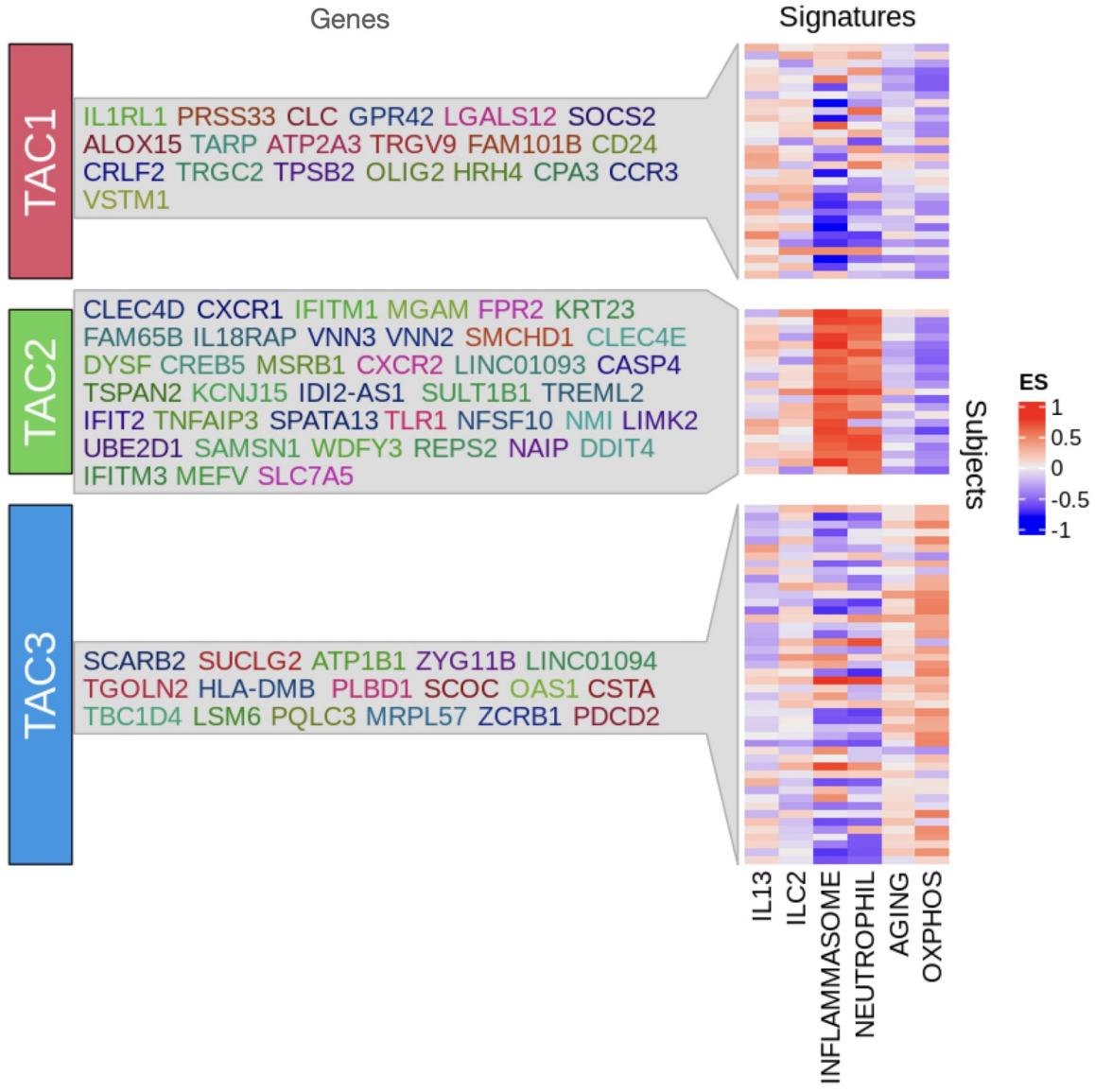
Both GSEA and GSVA was used during the analysis within this thesis to calculate group-phenotype wise enrichment (GSEA) and sample-wise enrichment (GSVA) of a set of genes (gene signatures) in the sputum transcriptomics. Each gene signature utilised describes a set of genes associated with a biological pathway and were collected from the literature.

All gene signatures are listed in Appendix I. and each gene signature is referred by referencing the original literature from which the signature was derived from as well as the row on the Appendix Table (e.g. **Sig1** (189)) . In addition both the KEGG (Kyoto encyclopaedia of genes and genomes) (190) and Gene Ontology (GO) (191) biological process database were used throughout for GSEA; where these are used they are referenced in the methods section of each results chapter.

2.2.8 Derivation of transcriptomic-associated clusters (TACs)

As previously described by Kuo *et al.* in 2017 (20), three transcriptome-associated clusters have been defined. Affymetrix microarray (GPL570 Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array) (Affymetrix) microarrays with RNA extracted from sputum cells (GSE76262) were used for gene expression profiling. Derivation of the three TACs was undertaken by applying hierarchical clustering based on the Euclidian distance of a reduced set of 508 genes that were differentially expressed between eosinophil-high and eosinophil-low subjects using sputum transcriptomics (20). The optimum number of clusters was established via consensus clustering, where the optimum number k is chosen when the consensus matrix histogram reaches a bimodal distribution (20). Gene signatures for each TAC were determined based on the nearest shrunken centroid (**Figure 12**).

Figure 12. TAC (transcriptome-associated clusters) definition by genes associated with each TAC group and enrichment of the pathways IL12, ILC2, Inflammasome, Neutrophil, Aging & Oxphos for each TAC group.



Legend:
Left: List of gene signatures associated with each of the TACs dervived from nearest shrunken centroid.

Right: GSVA heatmap of signatures associated with specific pathways. Heatmap showing the GSVA enrichment score of the pathway signatures for IL13-T2 [**SIG4**] (192), ILC2 [**SIG3**] (193), inflammasome [**SIG11**] (21), neutrophil [**SIG2**] (194), aging [**SIG10**] (195), and Oxidative Phosphorylation [**SIG5**] (196) grouped by the TACs. Data is taken from Kuo et al (20).

Chapter 3. Evaluation of metagenomic pipelines for U-BIOPRED metagenomic sequence data

Parts of this chapter are based on the publication:

***A Versi**, I Fransiskus Xaverius et al. H influenzae & M catarrhalis in sputum of severe asthma with inflammasome and neutrophil activation. Allergy 2023; 00:1–14.*

3.1 Introduction

Metagenomics has been a recent development in the field of microbiology and genomics that allowed scientists to explore the genomic and genetic diversity of microorganisms and entire microbial communities found in environmental and clinical samples. Metagenomics can be divided into two distinct approaches; amplicon metagenomics, a targeted approach that focuses on specific genes or markers for taxonomic classification (including 16S/18S/26S rRNA and intergenic transcribed spacer (197,198), and shotgun metagenomics, which provides a more comprehensive and unbiased view of both taxonomic and functional aspects of microbial communities based on all DNA sequences in the samples (199,200).

This powerful technique has the potential to transform current understanding of microbial ecosystems, enabling researchers to predict major metabolisms in surveyed environments as well as uncover novel microorganisms, study their functional potential, and investigate their roles in shaping the environment and human health (201–203).

3.1.1 Shotgun metagenomics

Metagenomics generally follows a four-step sequential process: 1) sampling and DNA extraction, 2) sequencing, 3) bioinformatics processing, and 4) analysis and visualisation. One of the key methodologies in metagenomics is shotgun metagenomic sequencing, which is a method of randomly fragmenting all DNA present in a given sample, followed by the sequencing of these fragments. Shotgun metagenomics addresses the limitations of amplicon sequencing, which targets specific genetic markers (e.g. 16S rRNA genes); shotgun metagenomics provides a

more comprehensive view of the entire genetic material within a sample by sequencing all DNA fragments and allowing the recovery of the entire length of target genes. It yields vast amounts of sequence data, including the DNA of known and unknown microorganisms, viruses, and other genetic elements. That being said, both short-read and long-read sequencing techniques for gene assembly face limitations. For example, the presence of universal and highly conserved regions complicates the assembly of short-read sequencing. Moreover, the long-read sequencing techniques used to overcome this limitation are typically plagued by high cost and unstable sequencing quality (204).

3.1.2 Bioinformatics pipeline for taxonomy classification and functional genomics

The vast amount of data generated by shotgun metagenomics requires sophisticated bioinformatics tools and pipelines for analysis (205,206). As with traditional microbiological studies, identifying the microbial taxon or taxa in the obtained DNA sequences is a critical step in the metagenomics analysis process. Taxonomy assignment can be performed using sophisticated classifiers and involves classifying each sequence into taxonomic groups, typically at various levels such as species, genus, family, and so on (207). This information is crucial for understanding the composition and diversity of the microbial community within a sample.

3.1.3 Kraken, Kaiju and MetaPhlAn: Three metagenomic pipelines

Several bioinformatics pipelines have been developed to perform taxonomy classification and gene microbial pathway classification from shotgun metagenomic data. Taxonomic classification tools are based on different types of data, utilising either DNA, protein or markers. Kraken, Kaiju, and MetaPhlan2 are three of the most widely used pipelines that involve unique features and algorithmic approaches.

3.1.3.1 Kraken

Kraken is an ultrafast and highly accurate metagenomic classification tool that is designed to assign taxonomy to DNA sequencing based on the alignment of k-mers. By using reference genomes, Kraken is able to construct a database of k-mers and the lowest common ancestor of all organisms with genomes in which each k-mer can be found. Kraken will then match k-mers in the metagenomic data to the database in order to identify the most likely taxonomic origin of each sequence (208). Kraken allows for rapid sequence classification with a high sensitivity and specification. However, the database size of Kraken can become high, and advanced computational resources may be needed for running the program (207). MiniKraken is a variant of Kraken with a reduced memory requirement from 70GB of RAM to 4GB of RAM (208). Kraken 2 also aimed to address the limitations of large memory requirements with Kraken, which significantly reduced the memory needed without sacrificing the high accuracy of the program. Moreover, Kraken 2 offers a 5-fold higher speed compared to Kraken 1 and improved sensitivity in viral metagenomics analysis through its new translated search mode (209). Braken (Bayesian Reestimation of Abundance with Kraken) uses a Bayesian formula to perform the estimation of species and genus-level abundance from the Kraken classification results (210).

3.1.3.2 Kaiju

Kaiju is a taxonomic classification program that is based on protein-level sequence classification as opposed to relying on k-mers. Kaiju finds maximum (in-)exact matches of amino acid sequences on the protein-level using the Burrows–Wheeler transform (211). Kaiju relies on the alignment of metagenomic reads to a pre-processed database of annotated proteins from microbial reference genomes, thus leading to the identification of the taxon or taxa on the best matches. Kaiju offers important advantages, namely high speed, efficiency, sensitivity and precision, with reduced database size compared to Kraken (211), although its sensitivity remains database-dependent.

3.1.3.3 MetaPhlAn2

MetaPhlAn2 is a marker-based taxonomic classification tool distinct from Kraken and Kaiju, which rely on DNA and proteins references to assign taxa, respectively. Using clade-specific markers, MetaPhlAn2 unequivocally assigns reads from shotgun metagenomics to taxonomic clades from a database (212). In addition to taxonomic assignment, MetaPhlAn2 can provide an accurate estimation of the coverage, presence and abundance of markers and by extension, the organisms. strain-level analyses are also possible via sub-species markers (212). In addition to its high resolution in species-level profiling, MetaPhlAn2 is not computationally demanding, with one of the lowest memory requirements of all taxonomic classification tools (207). However, MetaPhlAn2 is not without its disadvantages; it is limited to species-level classification and may not capture the full taxonomic diversity of a studied sample. In fact, MetaPhlAn2 was one of the taxonomic classification tools with the lowest performance due to limited marker-based databases (207).

3.1.3.4 HUMAnN2

HUMAnN2 is a functional metagenomic profiling pipeline originating from the Huttenhower Lab (213), the same group which develops and maintains MetaPhlAn2. Using shotgun metagenomics and metatranscriptomics, HUMAnN2 follows a three-tier search approach to profile the metabolic pathway content contributed by known and uncharacterized species within a given sample. In the first tier, a custom gene database is constructed based on species found within the sample using functionally-annotated pangenomes. In the second tier, gene alignment statistics are generated per species in addition to the unmapped reads through the alignment of the entire sample to the constructed custom gene database (from the first tier). In the final and third tier, unmapped reads are aligned to databases such as UniProt (214). This is usually followed by the construction and quantification of metabolic pathway prevalence, achieved by stratifying the alignment statistics into per-species and unclassified contributions and gene abundance values.

In this chapter these pipelines were evaluated to decide which was the most suitable pipeline to use with the U-BIOPRED data going forward.

3.2 Methods

3.2.1 Participants:

As detailed earlier in Chapter 2 (Datasets and Methodology), the U-BIOPRED cohort consists of four groups. Severe non-smoking asthmatics (SAn), severe smoking or ex-smokers asthmatics (SAs/ex), mild-moderate asthmatics (MMA) and heathy controls (HC). Induced sputum samples were taken at a baseline time-point (BL) and, in a subset of patients with severe asthma only, at a longitudinal time-point (LT) 12 months later.

3.2.2 Sample collection, Library preparation and Sequencing:

Sputum induction was undertaken as described earlier in the Methods chapter. Following inhalation of hypertonic saline, induced sputum samples were collected from participants. Sputum plugs were isolated and frozen immediately at -70°C.

Samples were then sent to Second Genome (San Francisco, California, USA) for library preparation and metagenomic sequencing. Genomic DNA was extracted from the frozen sputum samples using the MoBio Tissue and Cells DNA Isolation Kit (Qiagen, Germantown, USA) at 250µl input volume. The DNA extraction was performed in a BioSafety Cabinet in an ISO8 Cleanroom. Each batch included negative and positive controls. Extracted DNA was immediately stored in a -20°C freezer.

For the metagenomic analysis, samples were prepared for sequencing with the Illumina Nextera kit (Illumina, San Diego, USA) and quantified with Quant-iT dsDNA High Sensitivity assays (ThermoFisher Scientific, Waltham, USA). Libraries were pooled and run with 100 base-pair paired-end sequencing protocols on the Illumina HiSeq 2500 platform.

Although Blank DNA extractions were performed on sterile *phosphate-buffered saline* (PBS), no analysis was performed to assess the degree of background contamination associated with DNA extraction and sequencing reagents, as the sequence files of the blanks were not returned by the sequencing company.

3.2.3 Quality control and host reads removal:

In order to check the quality of the metagenomic sequencing reads FastQC version 0.11.8 (215) was used. Then, bowtie2 version 2.3.5.1 (with its default parameters) (216) was used to align the quality-filtered reads to the human reference genome (hg38 [https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/]). Unmapped, non-host reads were separated from the host reads and sorted using samtools version 1.9 (217). The output files, which were in a BAM format, were converted into fasta files using bedtools version 2.28 (218) (for the conversion of BAM files into paired-end fastq files) and R/Bioconductor package ShortRead version 3.6.2 (219) (for the conversion of fastq files into fasta files).

3.2.4 Bioinformatics Pipelines:

Non-host metagenomic reads were subjected to bioinformatic pipelines to generate microbiome profiles. Three approaches were used to estimate microbiome profiles: Kraken2 with MiniKraken2 and Braken2 (208), Kaiju version 1.7.4 (211) and MetaPhlAn2 version 2.7.2 (220). Quality controlled metagenomic sequences were run through kraken2, kaiju and MetaPhlAn programmes using the following commands:

Kraken command:

```
Kraken2 --threads 32 --fastq-input --confidence 0.05
--bzip2-compressed {input_1.fastq.bz2} {input_2.fastq.bz2} --
output
{output.reads}) --report {output.report}
```

Kaiju command:

```
kaiju -t {db.nodes} -f {db.fmi} -i {input.1} -j {input.2} -o
{output}
-z 32
```

MetaPhlAn2 command:

```
metaphlan2.py --input_type fastq --nproc 32 --bowtie2out
{output.bowtie2out} {db} {input} {output.report}
```

The functional profiling that estimates the abundance of microbial gene families and pathways from metagenomic data was performed using the HUMAnN2 (213).

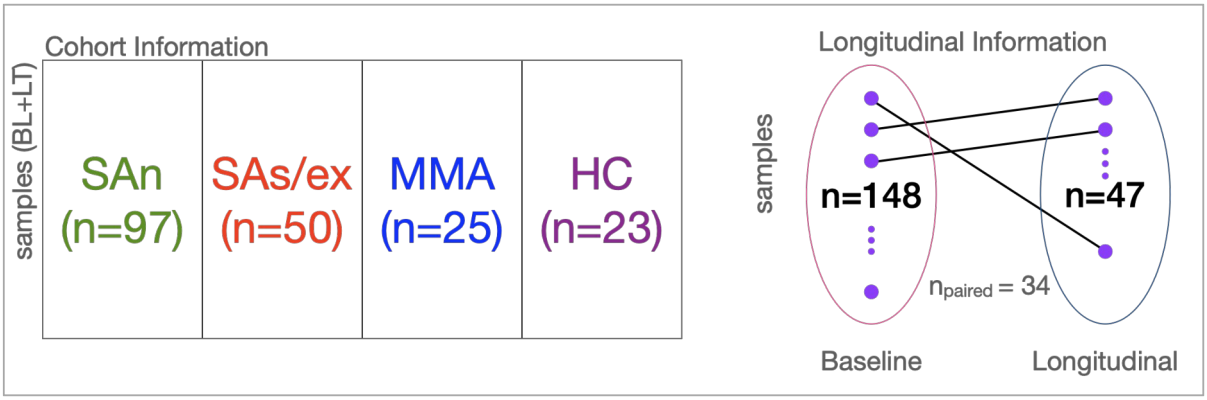
HUMAnN2 command:

```
humann2 --verbose --threads 4 --input {input} --output humann2
```

3.3 Results

195 paired read metagenomic sequence files were received from Second Genome (San Francisco, California, USA) and had passed quality control from the 374 number of sputum samples taken (**see Chapter 2**). These included 148 baseline samples and 47 follow-up samples (**Figure 13**). All the sequence data have been uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project accession PRJNA946921.

Figure 13. Cohort information after metagenomic sequencing of sputum.



Legend: Table showing the number of metagenomic samples of all subjects in the study at both baseline and longitudinal timepoints by disease severity cohorts. The number of samples at the baseline and longitudinal are shown with 34 subjects having samples at both timepoints. Study cohorts: SAn (non-smokers with severe asthma); SAs/ex (smokers or ex-smokers with severe asthma), MMA (non-smokers with mild or moderate asthma) and HC (healthy controls).

Out of the 47 follow-up samples only 34 were paired to corresponding baseline sample taken from the subject, due to samples not passing quality control.

The 195 paired read sequence files were fed through the three pipelines using the commands described above in the methods section after host removal.

Table 2. Summary of the output of the three pipelines

	MetaPhlAn2	KrakEN2	Kaiju
# Species	320	4,525	116
# Samples	145	195	148
Species Density	100%	25%	100%

Legend: Table depicting at the baseline time-point, the number of species, the number of samples with one or more non-zero *species-level assignments* and the species density (the fraction of non-zero species in the output table) of the three pipelines tested; MetaPhlAn2, KrakEN2 and Kaiju.

Table 2 depicts the number of species, the number of samples with one or more non-zero species-level assignments and species density of each of the three pipelines. The species density represents the fraction of non-zero elements in the abundance matrix.

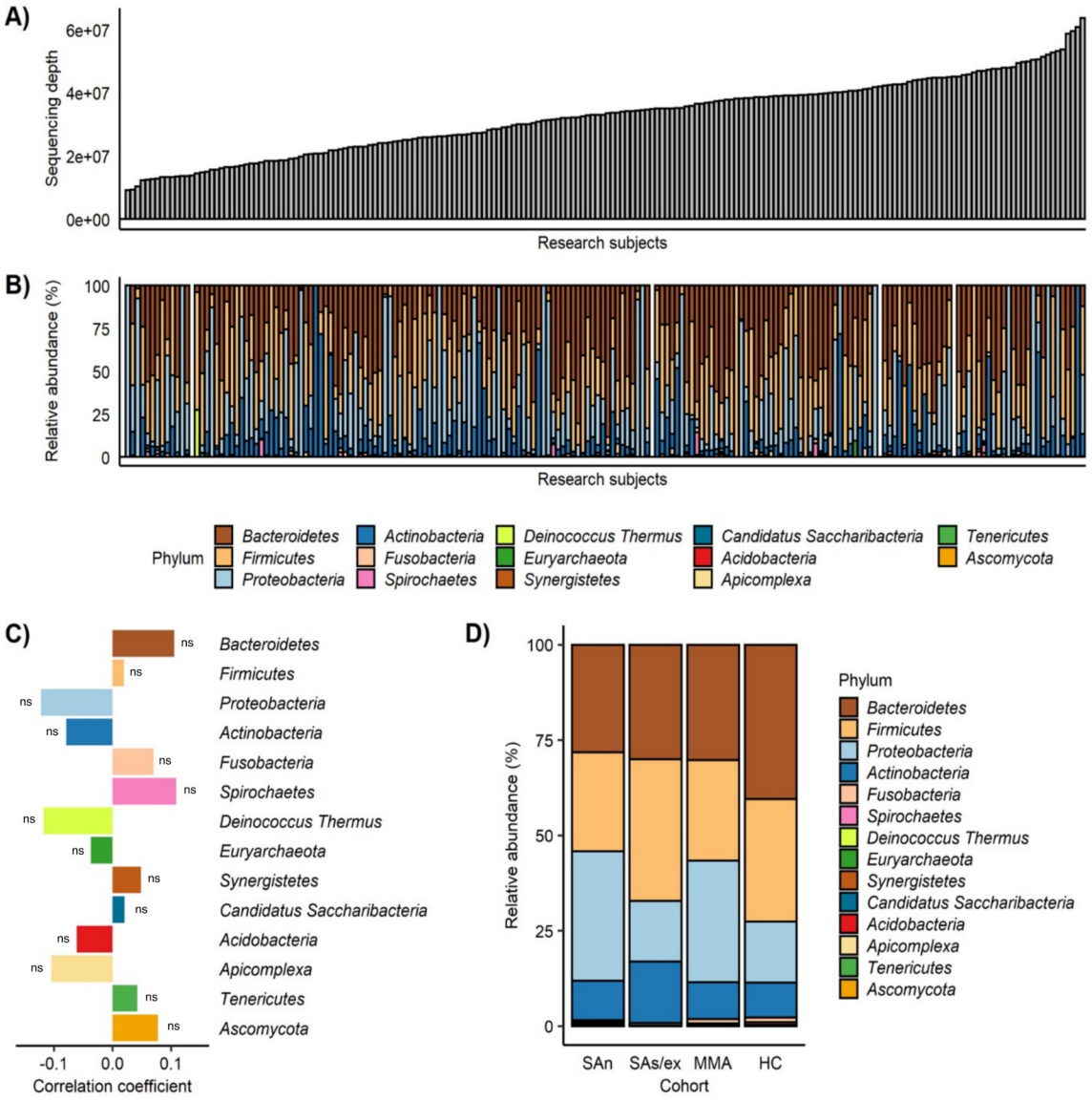
Unusual and unexpected species were found in both the Kraken and Kaiju output. The Kraken output consisted of dominant *Homo sapiens* even after host-read removal with bowtie2. Kaiju's output contained high abundances of the species *Ralstonia solanacearum*, *Klebsiella pneumoniae*, *Escherichia coli* and *Plasmodium ovale*. The MetaPhlAn2 output was more comparable with previously published data on the airway microbiome (85) and therefore the MetaPhlAn2 output was chosen to go forward with.

3.3.1 Microbiome profiling with MetaPhlAn2

Out of the 195 samples, 5 samples were excluded as their reads were not assigned to any phylum by MetaPhlAn2, leaving 190 samples**.**

A illustrates the depth of metagenomic sequencing of these 190 sputum samples before host read removal. The sequencing depth ranged from 9.1 million to 63.6 million reads, with a median and an interquartile range of 33.3 million and 17.2 million reads, respectively. The range of the depth of non-host sequencing reads had a median of ~1.5 million reads after removing host reads. The assessment of airway microbiome data was conducted at phylum level (**Figure 14B**). Airway microbiome profile was not affected by the sequencing depth (**Figure 14C**)**.** The overall abundance of bacteria was vastly (>95%) accounted for by four bacterial phyla, Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria (**Figure 14D**).

Figure 14. Assessment of metagenomic sequencing and overview of phylum-level microbiome profiles



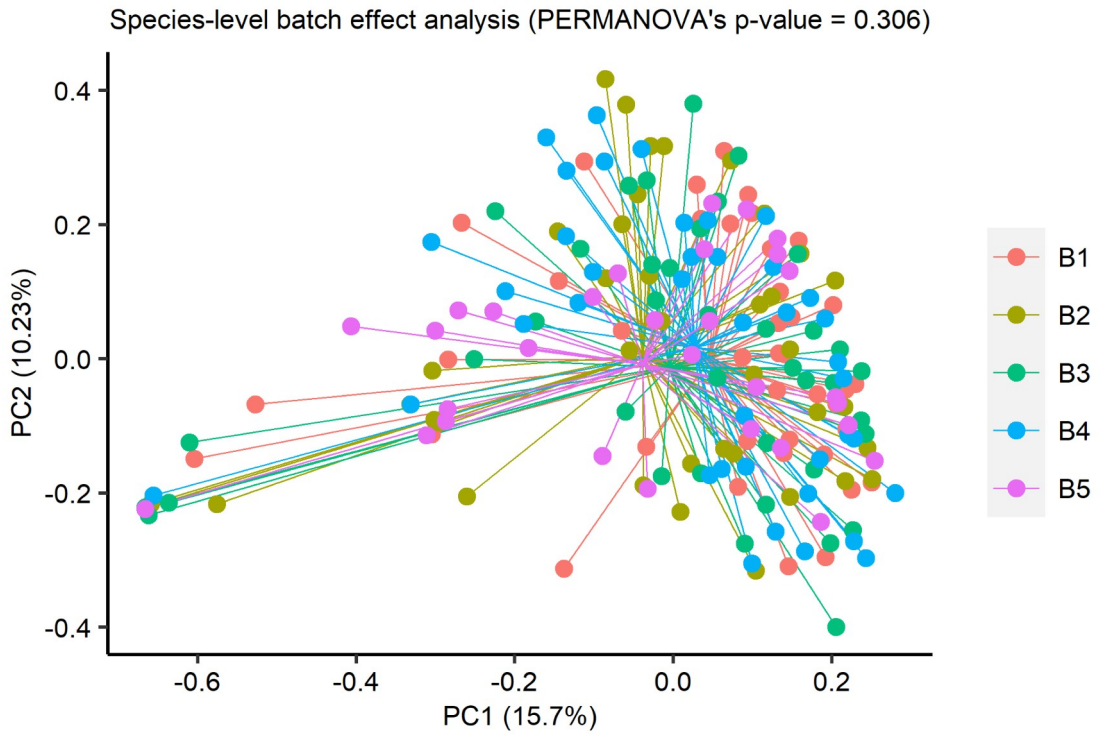
Legend: (A) Bar plot showing the metagenomic sequencing reads of sputum DNA of all subjects in the study ordered by increasing depth. (B) bagplot showing the phylum-level airway microbiome profiles of subjects in the study using the same order of subjects as in (A). (C) bar plot showing Spearman correlation between the relative abundance of bacterial phyla and the depth of metagenomic sequencing reads. (D) bagplot showing the average phylum-level airway microbiome profiles across study cohorts. Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001. Study cohorts: SAn (non-smokers with severe asthma); SAs/ex (smokers or ex-smokers with severe asthma), MMA (non-smokers with mild or moderate asthma) and HC (healthy controls).

After species-level reads assignment by MetaPhlAn2, 189 samples remained consisting of 91 samples for SAn, 50 for SAs/ex, 25 for MMA and 23 for HC. Both the SAn and SAs/ex cohorts contained samples from baseline and follow-up timepoints, with 97 (64 from SAn and 33 from SAs/ex) samples coming from the baseline

timepoint and 44 coming from the follow-up time-point (28 from SAn and 16 from SAs/ex). Out of these only 34 paired samples with both time-points remained.

Batch effects were analysed using permutational multivariate analysis of variance (PERMANOVA) of the Bray-Curtis dissimilarity of the species-level abundance data (**Figure 15**) which showed there was no significant difference between the composition of batches, indicating that contaminants were either low or consistent.

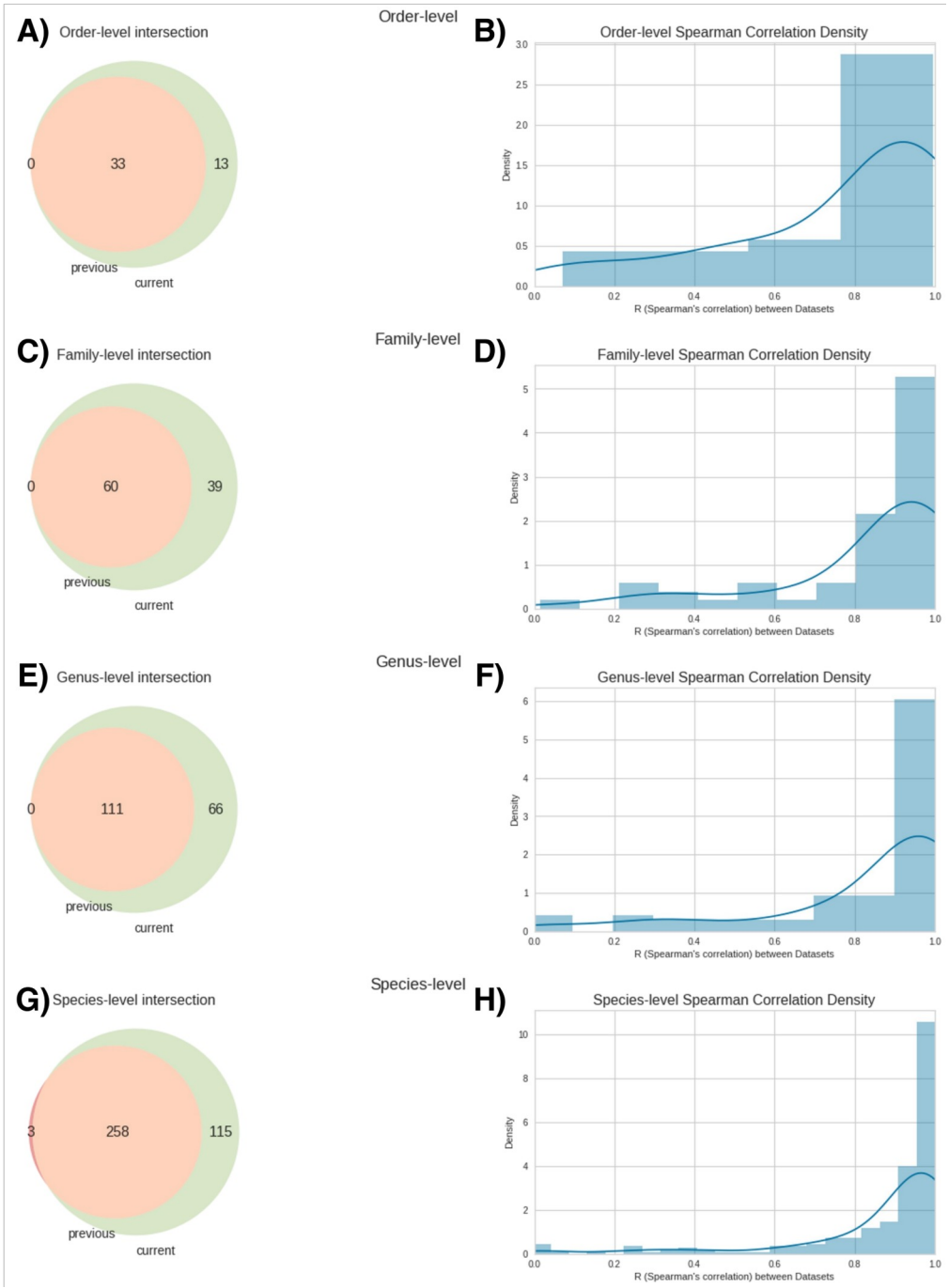
Figure 15. Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes across sample batches B1-B5.



Legend: Plot showing the principle coordinate analysis (PCoA) on the brays-Curtis dissimilarity matrix of the metagenomic data, where the colours represent each batch. PERMANOVA test showed a non-significant difference between the batches.

compares previously used and published taxonomic abundance data from an older version of the pipeline (87) run by Second Genome with the taxonomic abundance data generated from the updated version (2.7.2) of the pipeline. The figure shows an increase in all taxa, from order to species level and the abundance of all the common taxa from both versions of the pipeline are well correlated.

Figure 16. The Venn diagram depicts the overlap for the two datasets at each taxa-level.



Legend: (A) order-level, (C) family-level, (E) genus-level and (G) species-level. The density plots depict the frequency of the spearman's correlation value for each taxa at each taxa-level, (B) order-level, (D) family-level, (F) genus-level and (H) species-level. The figure depicts that the updated abundance dataset has picked up more taxa and the intersection between the two datasets are well correlated.

3.4 Discussion

For microbiome profiling MetaPhlAn2's results were more realistic than both the results of Kraken and Kaiju. This is because the top taxa of Kaiju's results were dominated by unexpected microbes including *Ralstonia*, *Klebsiella* and *Escherichia* (with the corresponding species are *Ralstonia solanacearum*, *Klebsiella pneumoniae* and *Escherichia coli*). MetaPhlAn2 implements an approach that aligns sequencing reads to marker genes. It is fast and requires very low memory, but its performance is usually limited by the bias introduced by the uneven distribution of sequence markers across microbes of interest (207). The Kraken output contained a high abundance of human reads even after human read removal with bowtie2. This may be due to MiniKraken's reduced k-mer alignment database which may have introduced more false positive assignments compared to MetaPhlAn2.

Due to the metagenomic sequencing of the U-BIOPRED samples occurring between 2010 and 2015, the standard read length at the time was a 100 base-pair (bp) paired-end sequencing protocol and hence the sequence data read length is shorter than the current standard of 150 bp or higher (221). This imposes a few limitations when it comes to pipeline choice. As noted in its original publication (211), Kaiju is shown to be much less sensitive for classifying reads whose length is approximately 100 bp (on average, the sensitivity for genus-level classification is about 40%), compared to reads whose read length is approximately 250 bp (sensitivity is about 70%). This short read length may also be the reason why the Kraken algortihm with the MiniKraken database pipeline produced abundance data containing unusual species such as *Homo sapiens* even after host removal. For short-reads the sensitivity of Kraken with MiniKraken can decrease by up to 11% (208).

MetaPhlAn2 is therefore a more preferrable pipeline although one of the consequences of this is that some samples were not mapped to known reference microbial markers and thus not included in further analysis.

Another limitation to the taxonomic classification and abundance data is the inherent biases generated from the Illumina HiSeq sequencing platform and how this may impact the interpretation of the abundance data. Several studies have described guanine-cytosine (GC) content bias, where there is an underrepresentation of GC

poor organisms (222,223). Another example is the higher sequencing yields of organisms with GC-balance across regions of *Plasmodium falciparum* (mean GC; 19%) and *Rhodobacter sphaeroides* (69%), as opposed to organisms with low-or high-content GC regions which had fewer yields (224).

In addition, another limitation is lack of sequenced negative control samples for each batch due to it not being kept by the sequencing company. This therefore means that batch specific contaminants cannot be shown and removed. However, **Figure 15**, demonstrates that differences in the taxonomic abundance between batches is non-significant and hence gives more confidence that the lack of accounting for negative controls in the downstream analysis do not significantly change the outcome between batches.

Having determined he optimal pipeline for analysis of U-BIOPRED sputum metagenomics, the examination of differential enrichment of microbial species across asthma severity and molecular phenotypes could proceed, which is the topic of the next chapter in the thesis.

Chapter 4. Haemophilus influenzae and Moraxella catarrhalis in the sputum of severe asthma associated with inflammasome and neutrophil activation

Parts of this chapter are based on the publication:

***A Versi**, I Fransiskus Xaverius et al. H influenzae & M catarrhalis in sputum of severe asthma with inflammasome and neutrophil activation. Allergy 2023; 00:1–14.*

4.1 Introduction

Severe asthma is a chronic inflammatory disease of the airways, with complex and heterogenous clinical presentation (15). Among severe asthma subtypes, severe eosinophilic asthma is the most established inflammatory phenotype and is characterized by an increased production of type 2 inflammatory markers, namely IL-4, IL- 5 and IL- 13 cytokines (16). Conversely, type 1 inflammation, activation of the inflammasome activation and subsequent sputum neutrophilia are characteristic of the neutrophilic inflammatory phenotype (20). However, the mechanism underlying neutrophil recruitment and activation has yet to be elucidated, despite evidence linking inflammasome activation and neutrophil airway recruitment in asthma (21). In fact, the interaction of airway microbiota with host immunity and inflammatory responses in the lung suggests the potential role of the lung microbiome in modulating clinical and inflammatory asthma phenotypes (67,68).

A comprehensive evaluation of asthmatic patients' microbiome composition is enabled by advanced culture-independent techniques, such as 16S ribosomal RNA microarray and PCR. By applying such techniques, it becomes possible to link changes in the airway microbiota to clinical asthma presentation. For example, an increase in the frequency of proteobacteria in the airway of asthma patients (56) has been associated with airway hyperresponsiveness and poorer asthma control (83). More specifically, patients with neutrophilic asthma were found to have reduced sputum bacterial diversity compared to other asthma phenotypes, as well as higher levels of Proteobacteria, including *Moraxella* and *Haemophilus* (particularly *Haemophilus influenzae*) (134,161). Studies of microbiome composition have also

shown associations between stable asthma and concentrations of bacteria such as *Streptococcus pneumoniae, Moraxella catarrhalis (75,91)* and *Haemophilus influenzae* (56,83,91).

The hypothesis of this chapter is distinct microbial species, such as *Haemophilus influenzae* and *Moraxella catarrhalis* are associated with specific inflammatory and molecular phenotypes in severe asthma.

The present chatpter thus aimed to comprehensively characterize and investigate the microbiome of asthmatic patients down to the species level in an unprecedented use of metagenomic whole genome sequencing approach of induced sputum. This approach provided valuable insights into the relationship between airway dysbiosis and airway inflammation through an extensive assessment of clinically relevant microbial species. Based on this, we examined the relationship between microbial species and relevant factors such as asthma severity and granulocytic inflammation. The study also undertook the first-ever examination of the association between bacterial species abundance and molecular asthma phenotypes, in an effort to clarify the potential connection between these species and the host immunological and inflammatory responses (20).

4.2 Methods

The participants and cohorts, the sputum induction and initial sample handling together with the TACs derivation are described in **Chapter 2. Datasets and Methodology**.

4.2.1 Quality control of metagenomic data and host reads removal

Quality check for the original sequencing reads (using FastQC version 0.11.817) was followed by the alignment of quality-filtered reads to the human reference genome (hg38) using the default parameters of bowtie2 version 2.3.5.1(216). Samtools version 1.9 was used to sort unmapped, non-host reads after their separation from host reads (217). Output file conversion was done as follows: bedtools version 2.28 (218) was used to convert BAM format output files in fastq files, which were then converted into fasta files using R/Bioconductor package ShortRead version 3.6.2.

(219). Microbiome profiles were generated by subjecting non-host metagenomic reads of each sputum sample to bioinformatics pipelines. Microbiome profile estimation was done using MetaPhlAn2 version 2.7.7 and its marker database (220), excluding viruses from the output.

4.2.2 Statistical and computational analysis

Shannon's index within sample diversity, α-diversity (225) and Bray–Curtis dissimilarity between microbiome sample pairs were calculated using the R languages' vegan package (226). Similarity between pairs of profiles was calculated as 1- day, where day is the associated Bray–Curtis dissimilarity. Sample to sample diversity and β-diversity were then evaluated using the generated dissimilarity matrices between cohorts and timepoints using the PERMANOVA implemented in vegan's ADONIS function (226). The dissimilarity matrices also served to visualize the relationship between sample groups using the principal coordinate analysis implemented in R's phyloseq package (227) for the identification of differentially abundant microbiome taxa, the R's DESeq2 package (185) with test and fitType parameters set to 'Wald' and 'parametric', and with pseudocount of 1 addition to species-level MetaPhlAn2 data, were used in case of a significant results after PERMANOVA analysis. Correction of Wald test's p- values for multiple testing was done using the Benjamini–Hochberg method.

Differences between groups (alpha diversity) was evaluated with the Kruskal–Wallis test. When significant differences were detected on the Kruskal-Wallis test, results were subjected to post hoc comparisons using the Mann–Whitney U-test with a Holm–Bonferroni correction. Wilcoxon signed-rank test was used to test differences in alpha diversity over time using paired baseline and longitudinal follow- up. All performed statistical tests were two-tailed. The relationship between microbiome abundance and clinical traits and inflammatory biomarkers were determined via Spearman correlations.

4.3 Results

Table 3 summarizes the clinical and physiological characteristics and **Table 4** and **Table 5** the asthma participants according to granulocytic inflammation status and to molecular phenotype, respectively.

Species-level analysis was conducted for 189 of 193 sputum samples using MetaPhlAn2. Where available, paired samples were used for baseline and longitudinal analyses. This yielded 33 subjects in the final analysis, despite the 46 sputum samples being analysed at follow-up. Analysis was also conducted according to several factors, namely granulocytic inflammation defined by the level of sputum eosinophils and neutrophils (**Table 4**) (169) and molecular phenotypes defined as transcriptomic- associated clusters (TACs) (**Table 5**) (20).

Table 3. Characteristics of participants

	Baseline					Longitudinal		
	SAn	SAs/ex	MMA	HC	P value*	SAn	SAs/ex	P value**
Subjects, n	67	33	24	23	ND	30	16	ND
Age (years)	54.1 (12.6)	53.7 (10.2)	40.4 (15.8)	39.3 (13.9)	<0.01	56.7 (11.5)	56.8 (7.8)	NS
Females, n (%)	40 (59.7)	18 (54.5)	11 (45.8)	7 (30.4)	NS	17 (56.7)	8 (50.0)	NS
BMI (SD)	28.3 (5.5)	29.5 (5.8)	25.6 (4.9)	25.5 (3.2)	<0.01	28.0 (5.2)	28.1 (4.6)	NS
Current Smoking, n (%)	0	11 (33.3)	0	0	<0.001	0	4 (25.0)	<0.001
Oral corticosteroid daily, n (%)	23 (34.3)	16 (48.5)	0	0	<0.001	9 (30.0)	8 (50.0)	NS
ACQ-5 score, (SD)	2.3 (1.3)	2.4 (1.2)	1.0 (0.7)	0.1 (0.2)	<0.001	2.0 (1.3)	2.1 (1.0)	NS
Severe exacerbations, per year, (SD)	1.8 (1.9)	1.8 (2.2)	0.1 (0.3)	0	<0.01	1.6 (1.7)	1.7 (2.4)	NS
Nasal polyposis, n (%)	24 (35.8)	9 (27.3)	4 (16.7)	2 (8.7)	NS	11 (36.7)	5 (31.2)	NS

Eczema, n (%)	18 (26.9)	8 (24.2)	9 (37.5)	0	NS	22 (73.3)	12 (75.0)	NS
Allergic rhinitis, n (%)	28 (41.8)	9 (27.3)	13 (54.2)	2 (8.7)	NS	12 (40.0)	4 (25.0)	NS
Non-Allergic rhinitis, n (%)	11 (16.4)	7 (21.2)	2 (8.3)	0	NS	9 (19.6)	6 (20.0)	NS
GERD, n (%)	26 (38.8)	18 (54.5)	3 (12.5)	1 (4.3)	<0.05	10 (33.3)	10 (62.5)	NS
Hay fever, n (%)	29 (43.3)	16 (48.5)	15 (62.5)	3 (13.0)	NS	10 (33.3)	5 (31.2)	NS
Blood Eosinophils μL^{-1}, mean (SD)	0.4 (0.3)	0.3 (0.2)	0.2 (0.1)	0.1 (0.1)	NS	0.4 (0.4)	0.3 (0.3	NS
Blood Neutrophils μL^{-1}, mean (SD)	5.2 (2.3)	5.8 (2.8)	3.6 (1.2)	3.5 (1.4)	<0.001	5.1 (2.0)	5.1 (2.6)	NS
Sputum Neutrophils (%), mean (SD)	58.8 (28.2)	56.1 (19.6)	48.7 (24.8)	42.6 (28.0)	NS	58.6 (28.2)	53.1 (19.5)	NS
Sputum Eosinophils (%), mean (SD)	14.0 (21.7)	12.5 (16.4)	4.0 (11.3)	0.3 (0.5)	<0.05	9.2 (17.1)	13.7 (17.6)	NS
FEV1% predicted, mean (SD)	62.2 (21.9)	67.6 (17.4)	90.5 (19.4)	104.1 (11.8)	<0.001	64.3 (19.5)	67.9 (19.1)	NS
FVC % predicted, mean (SD)	87.4 (18.6)	93.5 (19.0)	108.2 (17.0)	110.5 (11.1)	<0.05	89.7 (19.3)	95.1 (19.7)	NS
FEV1/FVC Ratio, mean (SD)	0.70 (0.16)	0.72 (0.12)	0.83 (0.12)	0.94 (0.07)	<0.001	0.71 (0.13)	0.71 (0.12)	NS

Abbreviations:

ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; HC: Healthy Control group; MMA: Mild-Moderate Asthma; N/A: Not applicable; ND: Not done; NS: Not significant; SAn: Severe Asthmatic non-smoking; SAs/ex: Severe Asthmatic smoking or ex-smoker; SD: Standard Deviation.

P value*: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the three asthmatic cohorts: SAn, SAS/ex, MMA.

P value**: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the two severe asthmatic cohorts: SAn, SAS/ex.

Table 4. Characteristics of subjects classified by sputum granulocytic inflammation type.

	EOS	NEU	MIX	PAU	P value*
Subjects, n	42	20	16	25	ND
Age (years)	52.8 (14.8)	49.0 (16.2)	53.8 (7.6)	49.0 (11.7)	NS
Females, n (%)	25 (59.5)	11 (55.0)	12 (75.0)	11 (44.0)	NS
BMI (SD)	27.2 (4.9)	29.1 (6.1)	27.5 (4.4)	28.1 (5.7)	NS
Current Smoking, n (%)	5 (11.9)	1 (5.0)		1 (4.0)	NS
Oral corticosteroid daily, n (%)	18 (42.9)	6 (30.0)	8 (50.0)	2 (8.0)	NS
ACQ-5 score, (SD)	2.0 (1.2)	2.2 (1.3)	2.1 (1.2)	1.6 (1.5)	NS
Exacerbations, per year, (SD)	2.0 (1.8)	2.6 (2.4)	2.8 (2.2)	1.1 (2.1)	<0.05
Nasal polyposis, n (%)	14 (33.3)	7 (35.0)	7 (43.8)	6 (24.0)	NS
Eczema, n (%)	9 (21.4)	11 (55.0)	6 (37.5)	6 (24.0)	NS
Allergic rhinitis, n (%)	26 (61.9)	8 (40.0)	8 (50.0)	10 (40.0)	NS
Non-Allergic rhinitis, n (%)	6 (14.3)	3 (15.0)	5 (31.2)	2 (8.0)	NS
GERD, n (%)	16 (38.1)	9 (45.0)	7 (43.8)	8 (32.0)	NS
Hay fever, n (%)	23 (54.8)	10 (50.0)	6 (37.5)	13 (52.0)	NS
Blood Eosinophils μL^{-1}, mean (SD)	0.4 (0.3)	0.2 (0.1)	0.4 (0.3)	0.2 (0.2)	<0.001
Blood Neutrophils μL^{-1}, mean	5.4 (2.4)	5.5 (3.1)	5.2 (1.7)	4.2 (1.5)	NS

(SD)					
Sputum Neutrophils (%), mean (SD)	42.5 (17.5)	86.8 (7.0)	84.5 (8.1)	42.1 (19.7)	<0.001
Sputum Eosinophils (%), mean (SD)	25.3 (22.1)	0.3 (0.3)	6.4 (6.0)	0.5 (0.5)	<0.001
FEV1% predicted, mean (SD)	62.3 (18.7)	70.0 (25.4)	56.3 (24.9)	80.5 (20.8)	<0.01
FVC % predicted, mean (SD)	91.2 (18.1)	88.5 (20.1)	88.1 (21.4)	101.9 (18.6)	NS

Abbreviations:

ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; EOS: Sputum eosinophilic high; NEU: Sputum neutrophilic high; MIX: Both Sputum eosinophilic and Sputum neutrophilic high; PAU: paucigranulocytic; N/A: Not applicable; ND: Not done; NS: Not significant; SD: Standard Deviation.

*: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the 4 inflammatory cohorts.

Table 5. **Characteristics of subjects classified by Transcriptome-Associated Cluster (TAC)**

	TAC1	TAC2	TAC3	P value*
Subjects, n	30	22	51	ND
Age (years)	56.47 (11.31)	49.55 (13.64)	49.02 (13.96)	<0.05
Females, n (%)	17 (56.67)	15 (68.18)	27 (52.94)	NS
BMI (SD)	27.28 (4.97)	28.54 (5.38)	27.88 (5.42)	NS
Current Smoking, n (%)	2 (6.67)	1 (4.55)	4 (7.84)	NS
Oral corticosteroid daily, n (%)	14 (46.67)	8 (36.36)	12 (23.53)	<0.05
ACQ-5 score, (SD)	2.05 (1.12)	2.27 (1.29)	1.80 (1.37)	NS
Severe exacerbations, per year, (SD)	2.17 (2.20)	1.41 (1.65)	1.18 (1.94)	NS
Nasal polyposis, n (%)	14 (46.67)	6 (27.27)	14 (27.45)	NS
Eczema, n (%)	14 (46.67)	6 (27.27)	14 (27.45)	<0.05
Allergic rhinitis, n (%)	12 (40.00)	10 (45.45)	20 (39.22)	NS
Non-Allergic rhinitis, n (%)	6 (20.00)	6 (27.27)	4 (7.84)	NS
GERD, n (%)	0 (33.33)	10 (45.45)	20 (39.22)	NS
Hay fever, n (%)	16 (53.33)	9 (40.91)	27 (52.94)	NS
Blood Eosinophils μL^{-1}, mean (SD)	0.47 (0.35)	0.28 (0.20)	0.26 (0.24)	<0.01
Blood Neutrophils μL^{-1}, mean (SD)	5.77 (2.45)	5.58 (2.78)	4.49 (1.85)	<0.05

Sputum Neutrophils (%), mean (SD)	48.23 (25.58)	88.04 (8.86)	49.78 (20.59)	<0.001
Sputum Eosinophils (%), mean (SD)	32.90 (22.00)	1.70 (2.30)	3.08 (4.28)	<0.001
FEV1% predicted, mean (SD)	8.10 (23.95)	63.72 (22.89)	76.96 (18.76)	<0.001
FVC % predicted, mean (SD)	8.51 (20.03)	85.05 (20.30)	98.98 (17.35)	<0.01

Abbreviations:

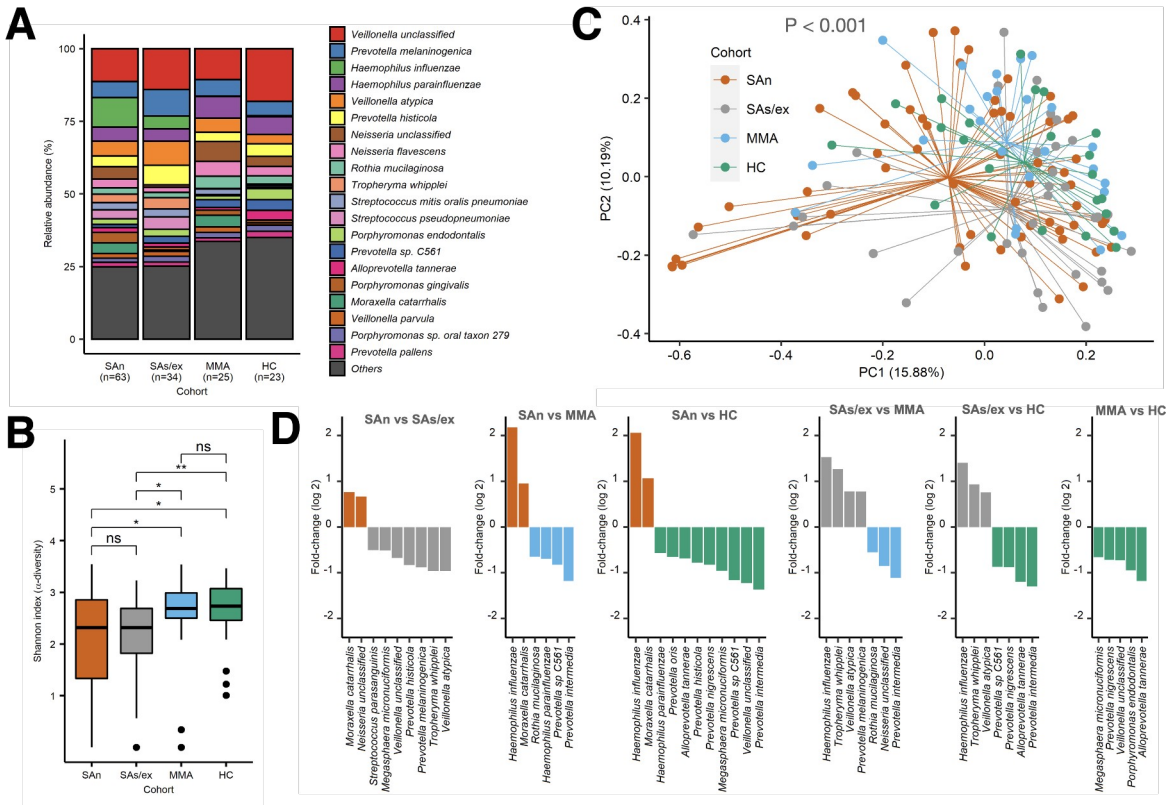
ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; HC: Healthy Control group; MMA: Mild-Moderate Asthma; N/A: Not applicable; ND: Not done; NS: Not significant; SAn: Severe Asthmatic non-smoking; SAs/ex: Severe Asthmatic smoking or ex-smoker; SD: Standard Deviation.

P value*: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the three TACs.

4.3.1 α- and β-diversity

Figure 17A shows the relative abundance of bacterial species by asthma severity. At species level, study cohorts exhibited differences in airway microbiome in both α-diversity ($p = .001$, Figure 17**B**) and β-diversity ($p \leq .002$, Figure 17**C**). Severe asthmatics had similar levels of α-diversity, regardless of smoking status (SAn and SAs/ex). However, patients in the MMA and HC cohorts have higher α-diversity when compared to SAs/ex (Figure 17**B**).

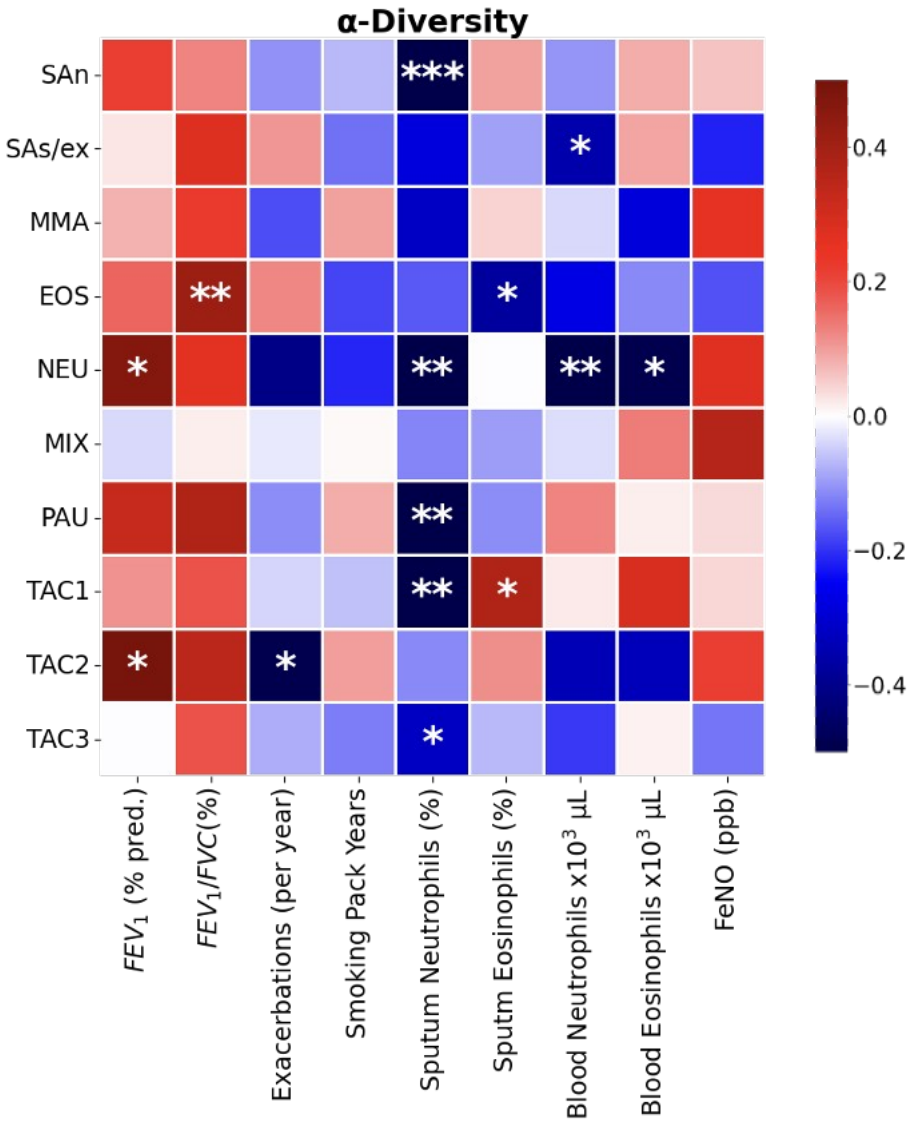
Figure 17. Airway of severe asthma subjects exhibits distinct species-level microbiome profiles



Legend: (A) Barplot showing the average species-level airway microbiome profiles across study cohorts. (B) Boxplot illustrating species-level α-diversity of airway microbiome across study cohorts. (C) Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes in the severity cohorts. (D) Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between the 4 groups of SAn, SAs/ex, MMA and HC. Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001. Study cohorts: SAn (non-smokers with severe asthma); SAs/ex (smokers or ex-smokers with severe asthma), MMA (non-smokers with mild or moderate asthma) and HC (healthy controls).

An inverse correlation was found between α- diversity and sputum neutrophils in SAn, neutrophilic and paucigranulocytic inflammation groups, and in TAC1 and TAC3 (Figure 18). Number of exacerbations in the previous year or smoking pack-years were not correlated with α-diversity, as opposed to FEV1 (% predicted) which was positively correlated to α- diversity in neutrophilic inflammation and in TAC2 (Figure 18).

Figure 18. Heatmap of Spearman's correlation coefficients



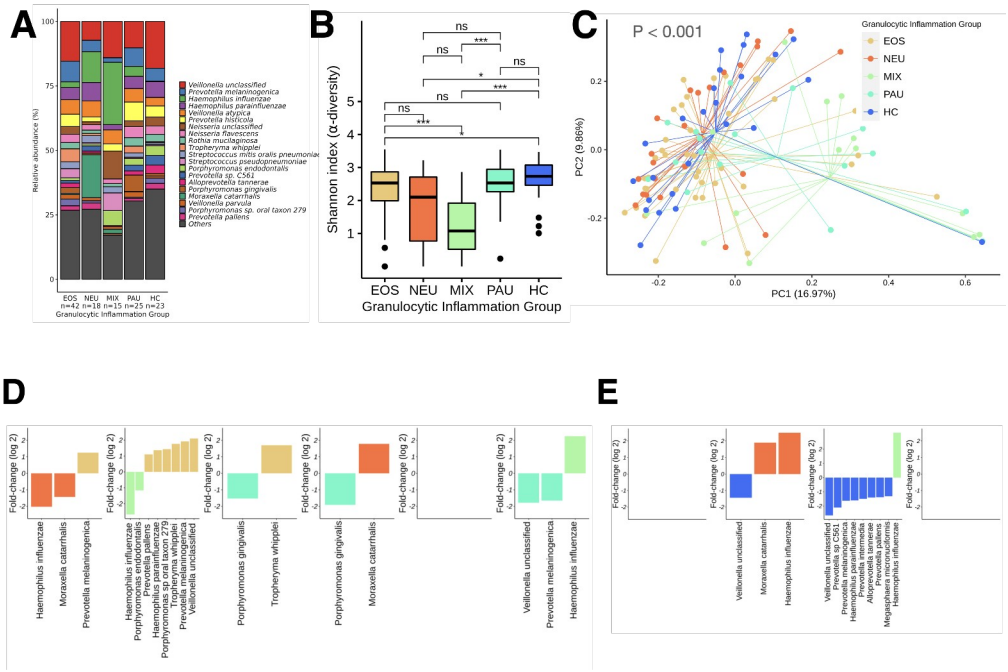
Legend: The heatmap depicts the spearman's correlation between FEV1 (% predicted), FEV1/FVC ratio, exacerbations in previous year, sputum neutrophil and eosinophil cell counts (%), blood neutrophil and eosinophil counts, fractional exhaled nitric oxide (FeNO) levels and smoking pack years against bacterial species α-diversity in the 3 classifications studied: asthma severity cohorts (San, SAs/ex and MMA), sputum inflammatory types (EOS, NEU, MIX and PAU) and transcription-associated clusters (TAC1, TAC2 and TAC3).

SAn: non-smokers with severe asthma; SAs/ex: smokers or ex-smokers with severe asthma; MMA: mild-moderate non-smoking asthma; EOS: eosinophilic high (sputum eosinophil % ≥ 1.49%); NEU: neutrophilic high (sputum neutrophil count % ≥ 73.6%); MIX: both neutrophilic and eosinophilic high; PAU: paucigranulocytic (sputum eosinophil % < 1.49% and sputum neutrophil count % < 73.6%); TAC: Transcription-Associated Cluster.

Significance of Spearman test for each group level correlation: not significant (adjusted p > 0.05) with clear boxes; *adjusted p ≤ 0.05; **adjusted p ≤ 0.01 and ***adjusted p≤0.001.

The species-level relative abundance across the granulocytic types are shown in Figure 19**A**. Lower α-diversity was observed across several inflammation phenotypes compared to HC, namely in eosinophilic ($p < .05$), neutrophilic ($p < .05$) and mixed granulocytic ($p < .001$) (Figure 19**B**). As for β-diversity, Bray–Curtis-based dissimilarity was evident across the inflammatory groups using PERMANOVA ($p < .001$) and ordination using principal coordinate analysis (Figure 19**C**).

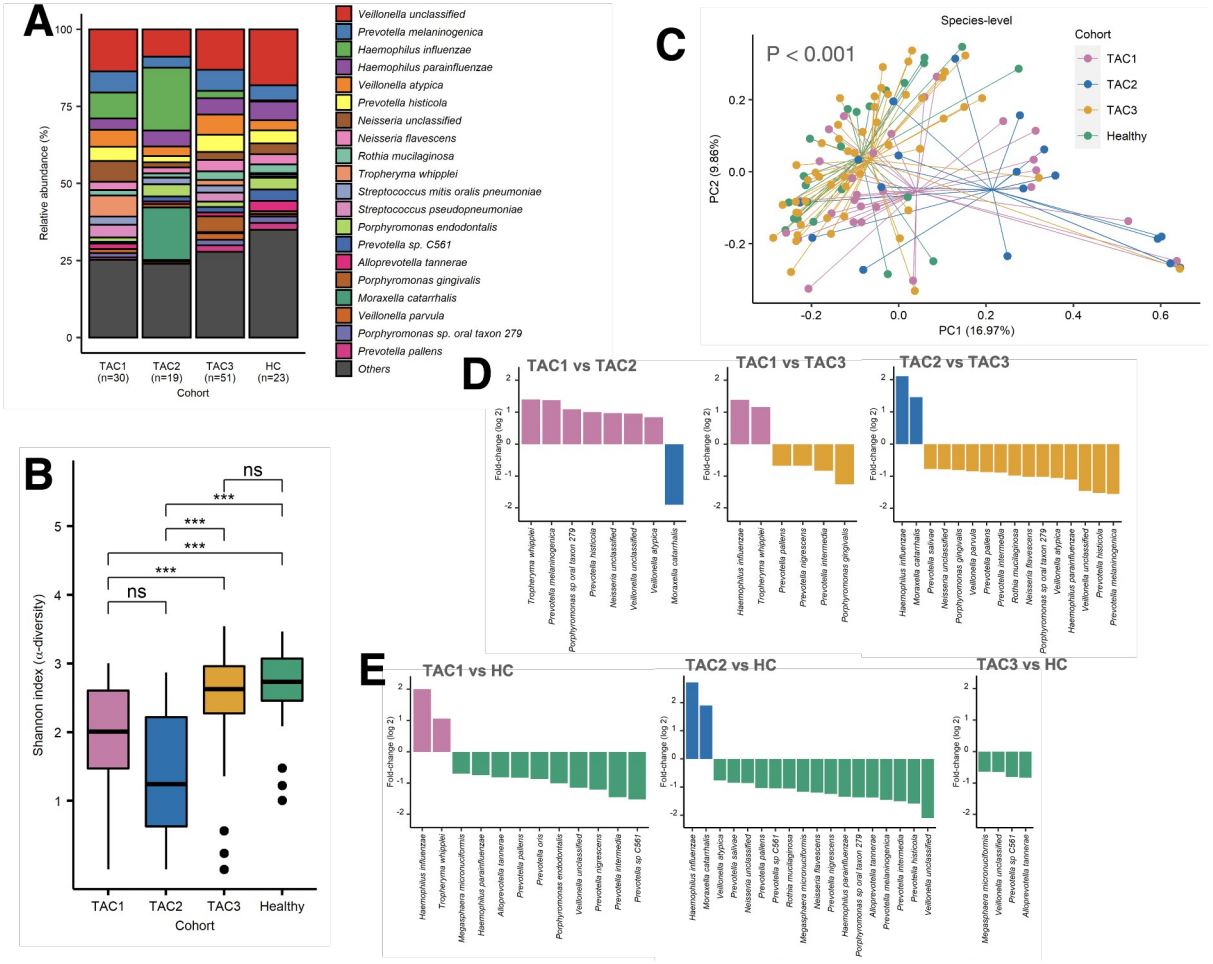
Figure 19. Sputum inflammation type shows distinct species-level microbiome profiles



Legend: (A) Barplot showing the average species-level airway microbiome profiles according to granulocytic sputum inflammation type. (B) Boxplot illustrating species-level α-diversity of airway microbiome according to granulocytic sputum inflammation type. (C) Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes in the 4 granulocytic sputum inflammation types and HC. (D) Barplots showing log2 fold-change of significantly differentially abundant airway microbial species according to granulocytic sputum inflammation type. (E) Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between each of the sputum granulocytic types and healthy control. Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001. Granulocytic sputum inflammation type: EOS: eosinophilic high (sputum eosinophil % ≥ 1.49%); NEU: neutrophilic high (sputum neutrophil count % ≥ 73.6%); MIX: both neutrophilic and eosinophilic high; PAU: paucigranulocytic (sputum eosinophil % < 1.49% and sputum neutrophil count % < 73.6%) and HC (healthy controls).

With regards to the TAC molecular phenotypes, TAC1 and TAC2 had reduced α-diversity compared to TAC3 and HC ($p < .001$) (Figure 20**B**). Differences in β-diversity based on Bray–Curtis dissimilarity were also observed across the TACs ($p < .001$) (Figure 20**C**).

Figure 20. Molecular phenotypes of asthma exhibit distinct microbiome profiles

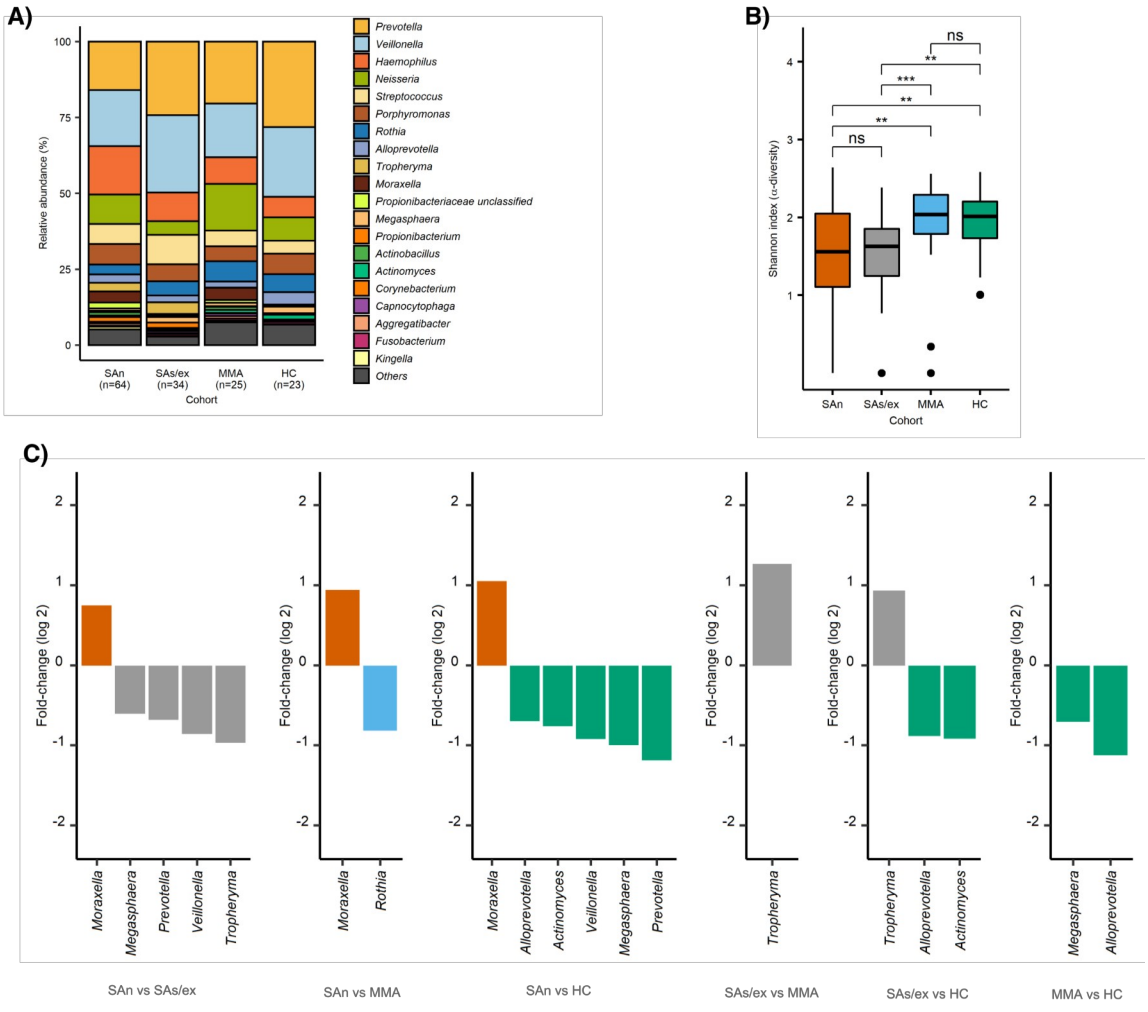


Legend: (A) Barplot showing the average species-level airway microbiome profiles across the 3 transcriptome-associated molecular phenotypes (TACs) and in healthy controls (HC). (B) Boxplot illustrating species-level α-diversity of airway microbiome across the TACs and HC. (C) Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes in the 3 TACs and HC (D) Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between the 3 TACs (E) Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between each TAC and HCs. Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001.

4.3.2 Differentially abundant genera and species by asthma grouping

Sputum samples from healthy subjects exhibited the following major genera: *Prevotella, Veillonella, Haemophilus, Neisseria, Streptococcus, Porphyromonas, Rothia, Alloprevotella* and *Tropheryma* (**Figure 21A**). Compared to HC, patients with severe asthma (SAn and SAs/ex) had higher abundance of *Moraxella* and *Tropheryma* (Figure 21**C**).

Figure 21. Airway of severe asthma subjects exhibits distinct genus-level microbiome profiles

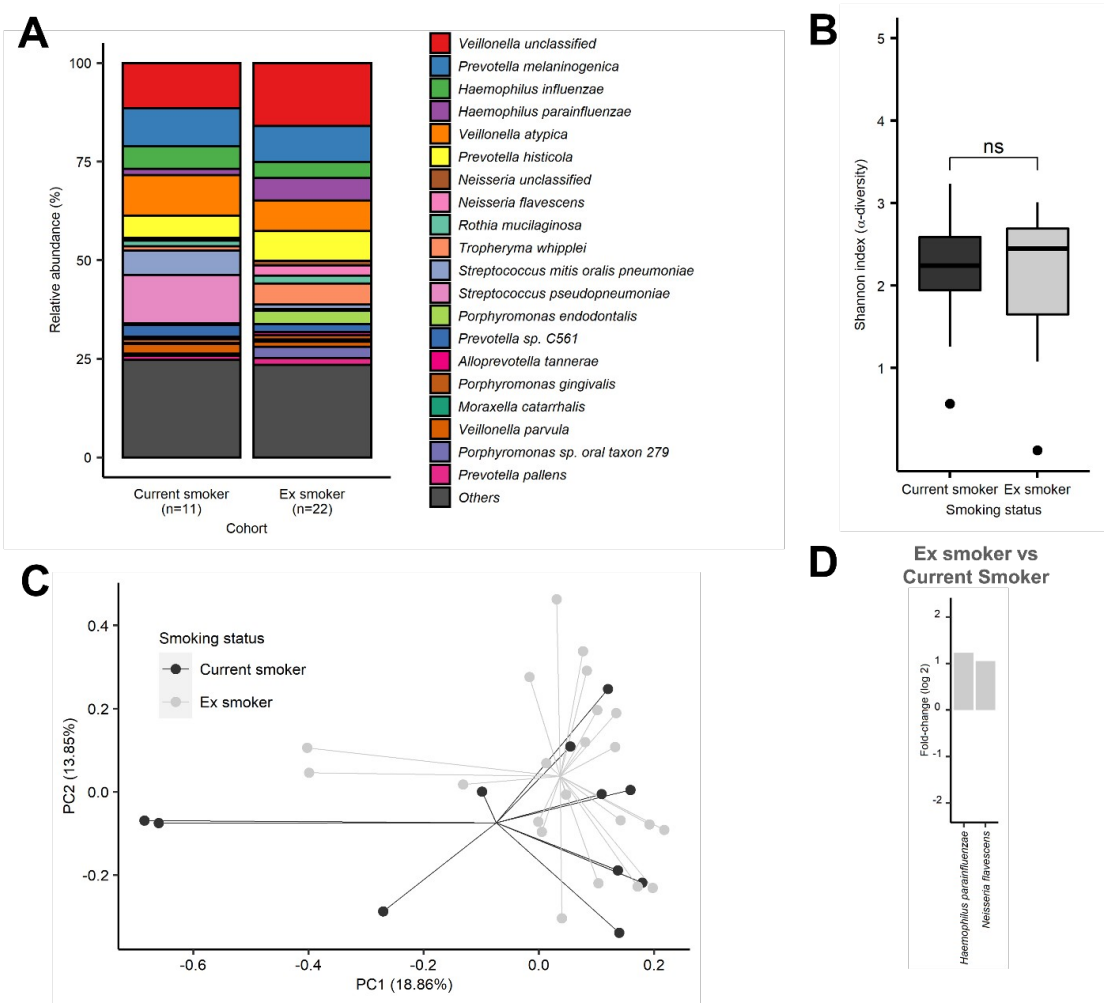


Legend: (A) Barplot showing the average genus-level airway microbiome profiles across study cohorts. (B) Boxplot illustrating genus-level α-diversity of airway microbiome across study cohorts. (C) Barplots showing log2 fold-change of significantly differentially abundant airway microbial genus between the 4 groups of SAn, SAs/ex, MMA and HC. Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001. Study cohorts: SAn (non-

smokers with severe asthma); SAs/ex (smokers or ex-smokers with severe asthma), MMA (non-smokers with mild or moderate asthma) and HC (healthy controls).

The most abundant microbes included *Haemophilus influenzae* and *Tropheryma whipplei* (Figure 17**A**). Compared to MMA and HC, non-smoking severe asthmatics (SAn) had higher abundance of *Haemophilus influenzae* and *Moraxella catarrhalis*. By contrast, previous or current smokers (SAs/ex) had higher levels of *Haemophilus influenzae, Tropheryma whipplei* and *Veillonella atypica* compared to both MMA and HC (Figure 17**D**). When examining the potential impact of smoking status (current vs ex-smoking), it was evident that *Haemophilus parainfluenzae* and *Neisseria flava* abundance was higher in ex-smokers compared to current smokers (Figure 22).

Figure 22. Barplots of the microbiome composition by smoking status

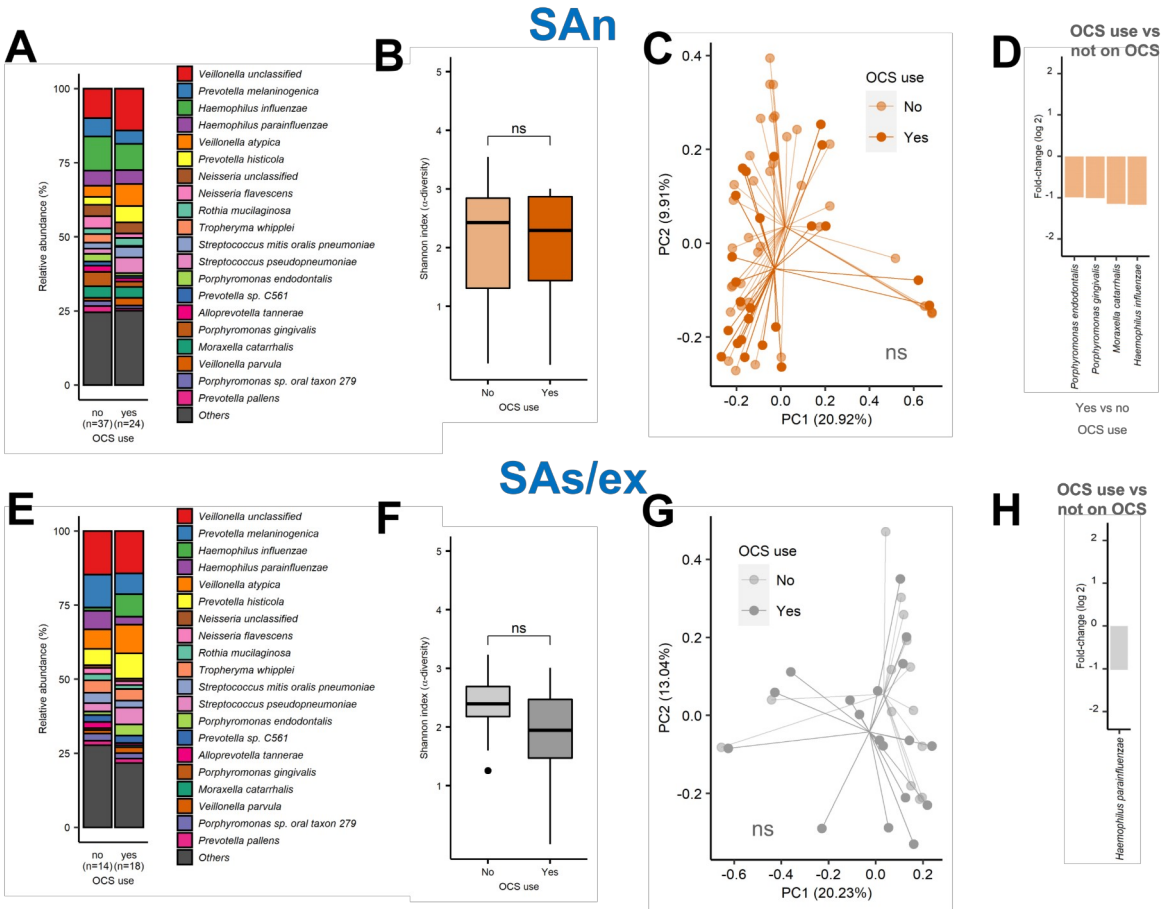


Legend: Microbiome composition of the average relative abundance of the top 20 species for the current and ex-smokers severe asthmatics (A). Boxplot illustrating species-level α-diversity of airway microbiome for the current and ex-smokers (FB). Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes in the severe asthma patients comparing ex-smokers against current-smokers (C). Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between ex-smokers and current-smokers in severe asthmatics (D). ns: non-significant.

As for the impact of exposure to daily OCS therapy, severe asthmatics (SAn and SAs/ex) had comparable α- and β-diversity whether they were on OCS or not (Figure 23).

On a species-level, *Haemophilus influenzae, Moraxella catarrhalis, Porphyromonas gingivalis* and *Porphyromonas endodontalis* had higher abundance in SAn that were not on OCS, while lower *Haemophilus parainfluenzae* abundance was observed in SAs/ex, (Figure 23).

Figure 23. Barplots of the microbiome composition by OCS usage

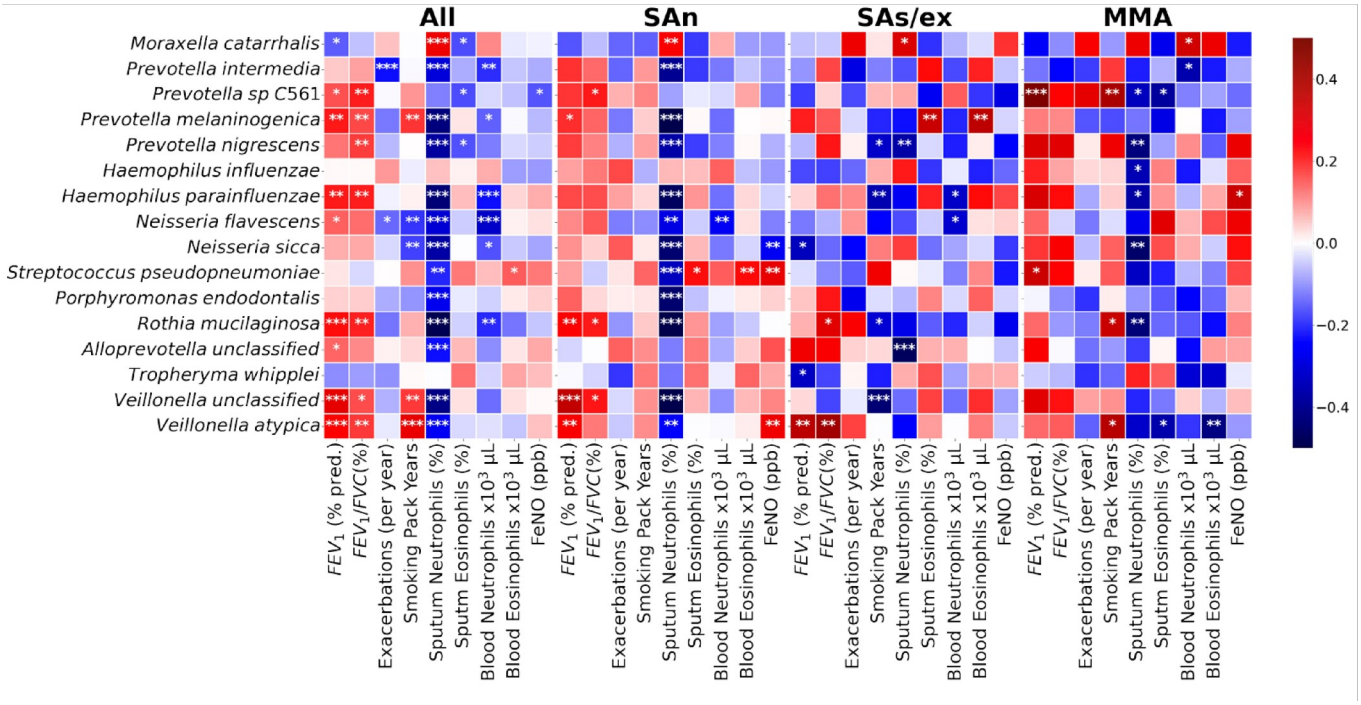


Legend: Microbiome composition of the average relative abundance of the top 20 species for severe asthma patients on daily oral corticosteroid (OCS) (yes) versus those not on daily OCS (no) in cohorts SAn (A-D) and SAs/ex (E-H). Boxplot illustrating species-level α-diversity of airway microbiome on daily oral corticosteroid (OCS) (yes) versus those not on daily OCS (no) in cohorts SAn (B) and SAs/ex (F). Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level alpha-diversity of airway microbiomes on daily oral corticosteroid (OCS) (yes) versus those not on daily OCS (no) in cohorts SAn (C) and SAs/ex (G). Barplots showing log2 fold-change of significantly differentially abundant airway microbial species between those on OCS and those not on OCS in cohorts SAn (D) and SAs/ex (H). ns: non-significant.

4.3.3 Species abundance and clinical and inflammatory features

The abundance of the top 16 species with clinical and inflammatory features in the whole group of asthmatics and in the 3 subgroups of SAn, SAs/ex and MMA were examined (Figure 24).

Figure 24. Heatmap of Spearman's correlation coefficients by species and severity



Legend: Spearman's correlations between FEV1 (% predicted), FEV1/FVC ratio, exacerbations in previous year, sputum neutrophil and eosinophil cell counts (%), blood neutrophil and eosinophil counts, and fractional exhaled nitric oxide (FeNO) levels against the abundance of top bacterial species for the 3 asthma cohorts. Significance of Spearman coefficient correlation: ns - not significant (adjusted p > 0.05), with clear box; *adjusted p ≤ 0.05; **adjusted p ≤ 0.01; ***adjusted p ≤ 0.001. SAn: non-smokers with severe asthma; SAs/ex: smokers or ex-smokers with severe asthma; MMA: mild-moderate non-smoking asthma; HC: healthy non-smoking and non-asthmatic.

4.3.3.1 Airflow obstruction

When evaluating indexes of airflow obstruction, a positive correlation was found between FEV1 and species including those of *Veillonella* and *Prevotella*, *Rothia mucilaginosa* and *Haemophilus parainfluenzae*. There was a positive correlation between *Prevotella intermedia, Prevotella melaninogenica, Veillonella unclassified* and *Veillonella atypica* in SAn, between *Veillonella atypica* and *Prevotella melaninogenica* in SAs/ex, and between Prevotella sp C561 in MMA with FEV1 (% predicted) **(Figure 24)**. In SAs/ex, there was a negative correlation of FEV1 (% predicted) and *Haemophilus influenzae, Neisseria sicca* and *Tropheryma whipplei*.

4.3.3.2 Exacerbations

In the entire study group, a negative association was observed between exacerbations in the previous year and *Prevotella melaninogenica* and *Prevotella intermedia*. However, the negative correlation with *Prevotella melaninogenica* emerged only in SAn (Figure 24). Meanwhile, *Neisseria sicca* was negatively correlated with exacerbations in the SAs/ex group, where a positive correlation also emerged with *Rothia mucilaginosa*.

4.3.3.3 Sputum neutrophils

Moraxella catarrhalis was positively correlated with sputum neutrophils but there was a strong negative correlation of sputum neutrophils with *Prevotella intermedia*, *Prevotella melaninogenica*, *Prevotella nigrescens*, *Alloprevotella unclassified*, *Haemophilus parainfluenzae*, *Neisseria flavescence*, *Neisseria sicca*, *Streptococcus pseudopneumoniae*, *Porphyromonas endodontalis*, *Rothia mucilaginosa*, *Veillonella unclassified* and *Veillonella atypica*. These correlations all persisted when examining the SAn group specifically, except for a lack of correlation with *Alloprevotella* unclassified. Correlations between sputum neutrophils and bacterial species in SAs/ex were negative for *Prevotella nigrescens* and *Alloprevotella unclassified*, but positive for *Moraxella catarrhalis* and *Neisseria sicca*. In MMA, *Prevotella sp C561*, *Prevotella nigrescens, Streptococcus pseudopneumoniae, Neisseria flavescens,*

Neisseria sicca, Rothia mucilaginosa, Haemophilus parainfluenzae and *Haemophilus influenzae* were negatively correlated with sputum neutrophils.

4.3.3.4 Sputum eosinophils

In regards to sputum eosinophils, *Moraxella catarrhalis, Prevotella sp C561*, and *Prevotella nigrescens* exhibited a negative association, as opposed to *Tropheryma whipplei* which was positively correlated to sputum eosinophils. Less associations were observed in the SAs/ex group, where sputum eosinophils were negatively correlated with *Rothia mucilaginosa*, but positively with *Tropheryma whipplei*. As for SAn, *Streptococcus pseudopneumoniae* and *Alloprevotella unclassified* showed positive associations (Figure 24).

4.3.4 Species abundance according to granulocytic inflammation

The abundance of *Haemophilus influenzae* and *Moraxella catarrhalis* was higher in the neutrophilic rather than the eosinophilic and paucigranulocytic groups, and more *Moraxella catarrhalis* compared to the mixed granulocytic group. Interestingly, *Tropheryma whipplei* was lower in the paucigranulocytic group and HC compared to the eosinophilic group, which also had more abundant *Tropheryma whipplei, Haemophilus parainfluenzae*, and other species including *Prevotella pallens, Prevotella melaninogenica, Porphyromonas sp oral taxon 279* and *Veillonella unclassified* compared to the mixed granulocytic group (**Figure 19C**).

4.3.5 Species abundance according to TAC molecular phenotypes

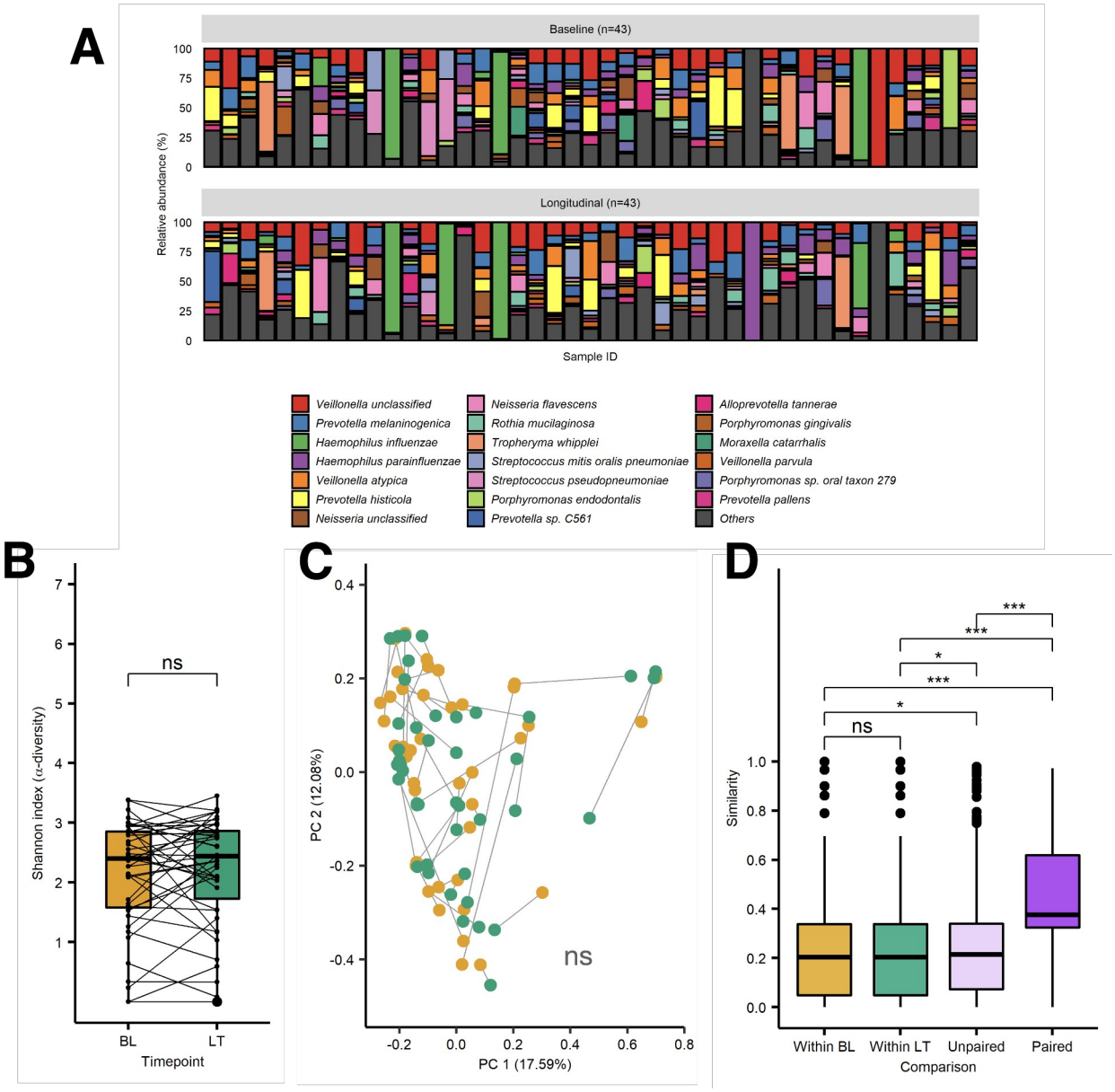
As previously described (18), TAC1 is characterized by the expression of immune receptors IL- 33R, CCR3R and TSLPR, TAC2 by TNF- and IFN-associated genes, and TAC3 by expression of high glucose and succinate metabolism genes **(Figure 24)**. There was an increase in relative abundance of *Tropheryma whipplei*, *Prevotella melaninogenica, Porphyromonas sp oral taxon 279*, *Prevotella histicola, Neisseria unclassified, Veillonella unclassified* and *Veillonella atypica* but decrease in relative abundance of *Moraxella catarrhalis* in TAC1 compared to TAC2. An increase in relative abundance of Haemophilus influenzae and *Tropheryma whipplei* in TAC1

compared to TAC3, and in *Haemophilus influenzae* and *Moraxella catarrhalis* in TAC2 compared to TAC3 (Figure 20**D**). It is worth mentioning that none of the samples in the TAC1 group exhibited *Moraxella catarrhalis*. By comparison to healthy controls, TAC1 which has the highest expression score for an IL-13 Th2 and an ILC-2 signatures had an excess of *Haemophilus influenzae* and *Tropheryma* whipplei while TAC2 which had the highest expression scores for inflammasome and neutrophil activation an excess of *Haemophilus influenzae* and *Moraxella catarrhalis*.

4.3.6 Longitudinal stability of microbiome according to asthma severity

Follow-up data from the 1-year visit of severe asthma patients participating in the study are shown in Table 3, when available. Overtime persistence of *Haemophilus influenzae* or *Tropheryma whipplei* dominance was observed: 3 out of 4 subjects with *Haemophilus influenzae* dominance at baseline and 2 out of 3 subjects with *Tropheryma whipplei* dominance at baseline maintained this prevalence at longitudinal follow- up (**Figure 25**).

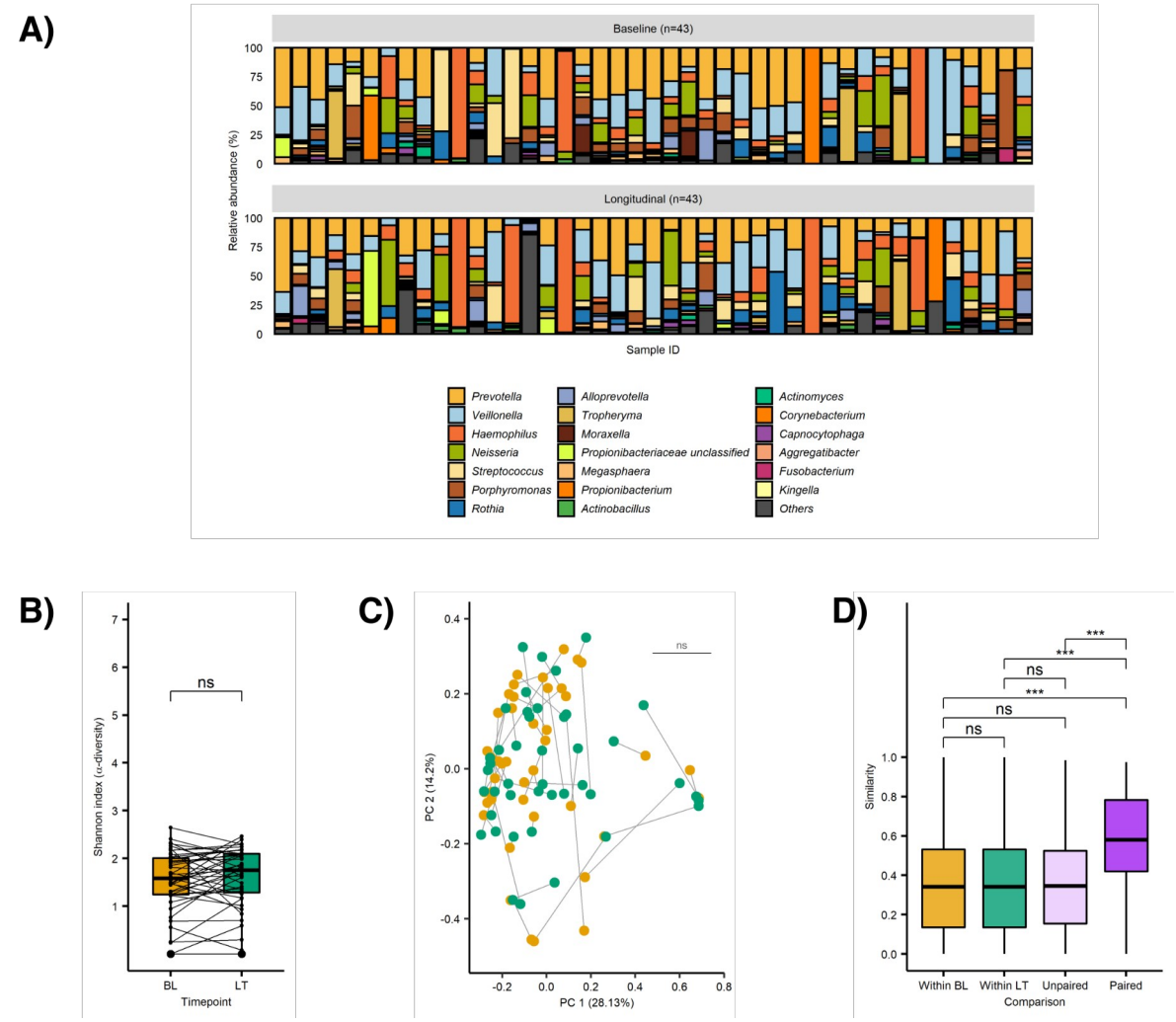
Figure 25. Airway of severe asthma subjects exhibits temporal stability in species-level microbiome profiles



Legend: (A) Barplots showing the baseline and longitudinal follow-up of species-level airway microbiome profiles in subjects with severe asthma, using the same order of subject identifiers. (B) Paired boxplot illustrating species-level α-diversity of airway microbiomes across in severe asthma cases. (C) Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating species-level β-diversity of airway microbiomes at baseline or longitudinal follow-up. Paired samples that are from the same subject are connected by a line. (D) Boxplot showing the species-level microbial compositional similarity across possible sample pairing: Within BL (pairing between any two baseline samples), Within LT (pairing between any two longitudinal follow-up samples), Unpaired (pairing between baseline and longitudinal follow-up samples from different subjects), and paired (pairing between baseline and longitudinal samples from the same subject). Timepoint: BL (baseline) and LT (longitudinal follow-up). Significance: ns (not significant, $P>0.05$), *$P\leq0.05$, **$P\leq0.01$ and ***$P\leq0.001$.

The same trend was observed at the genus level (Figure 26**A**). Longitudinal stability of airway microbiome was also observed in α- diversity (Figure 25**B**) and β- diversity (Figure 25**C**) of severe asthmatics at the species level. Similar results were seen at the genus level in α- diversity (Figure 26**B**) or β- diversity (Figure 26**C**). Moreover, Mann– Whitney U- test with Holm correction between paired and non- paired groups showed that genus- and species- level (Figure 25**D** and Figure 26**D**) microbiome compositional similarity between paired samples that were from the same subject was significantly higher compared to that from other possible sample pairings.

Figure 26. Airway of severe asthma subjects exhibits temporal stability in genus-level microbiome profiles



Legend: (A) Barplots showing the baseline and longitudinal follow-up of genus-level airway microbiome profiles in subjects with severe asthma, using the same order of subject identifiers. (B) Paired boxplot illustrating genus-level α-diversity of airway microbiomes across in severe asthma cases. (C) Principle coordinate analysis (PCoA) based on Bray-Curtis dissimilarity illustrating genus-level β-diversity of airway microbiomes at baseline or

longitudinal follow-up. Paired samples that are from the same subject are connected by a line. (D) Boxplot showing the genus-level microbial compositional similarity across possible sample pairing: Within BL (pairing between any two baseline samples), Within LT (pairing between any two longitudinal follow-up samples), Unpaired (pairing between baseline and longitudinal follow-up samples from different subjects), and Paired (pairing between baseline and longitudinal samples from the same subject). Timepoint: BL (baseline) and LT (longitudinal follow-up). Significance: ns (not significant, P>0.05), *P≤0.05, **P≤0.01 and ***P≤0.001.

4.4 Discussion

The results of this analysis sheds valuable insights on the role of microbial species in severe asthma. In this approach, we were able to examine differences in microbial species in the sputum of severe asthmatics from the level of asthma severity and inflammatory status to the extent of the molecular phenotype. This was achieved using metagenomic sequencing of samples from one of the largest cohorts of severe asthmatics.

Owing to their potential pathogenicity in respiratory diseases, three pathogenic microbial species were chosen as the focal point of the study. These were *Haemophilus influenzae*, *Moraxella catarrhalis* and *Tropheryma whipplei*. Analysis revealed reduced bacterial α- diversity of microbial species across both study groups. That being said, patients in the SAn cohort had higher frequency of *Haemophilus influenzae* and *Moraxella catarrhalis* compared to those in the MMA, HC, SA/ex groups. The latter (SA/ex) had a higher abundance of *Haemophilus influenzae* and *Tropheryma whipplei* compared to the MMA group only, with no other notable differences in microbial species between study groups. That being said, further examination of the microbiome with distinctions in regard to inflammatory status revealed the mixed granulocytic group as the one with the lowest α- diversity. While the neutrophilic group exhibited only a slight reduction in α- diversity, increased *Haemophilus influenzae* and *Moraxella catarrhalis* abundance was linked to neutrophilic inflammation. By contrast, increased levels of *Tropheryma whipplei* were associated with eosinophilic inflammation. The study also described the first attempt at assessing the relationship between microbial species and molecular phenotypes in asthma; TAC1 and TAC2 were associated with the highest abundance of *Haemophilus influenzae*. This was of interest considering that *Haemophilus influenzae* was highly present in both an eosinophilic (TAC1) and a neutrophilic

(TAC2) phenotype, but not a paucigranulocytic phenotype (TAC3). Other findings of note were a higher abundance of *Moraxella catarrhalis* in TAC2, but very low abundance in TAC1 and TAC3, while *Tropheryma whipplei* was increased in TAC1. The observed variation in *Haemophilus influenzae* abundance might be due to the reduced phagocytosis of this bacterium by lung macrophages in patients with severe asthma, leading to an increased abundance in the sputum (228). Current evidence suggests activation of toll-like receptor 2 (TLR2) (229) on human tracheal epithelial cells, leading to the release of inflammatory markers IL-1α, IL-1β, IL-6, IL-8, MCP-1 and TNF-α (230). While this cascade would result in neutrophil activation and inflammation, it is uncertain why *Haemophilus influenzae* is present in higher amounts in both TAC1 and TAC2, with no evident association to oral corticosteroid therapy.

Severe or poorly controlled asthma presenting with evidence of neutrophilic airway inflammation is known to occur as a result of respiratory infections caused by airway pathogens, one of which is *Moraxella catarrhalis (91,134)*. In the present study, sputum neutrophils were most positively correlated with *Moraxella catarrhalis* in both the SAn and Sas/ex groups. This was directly opposed to the MMA group, where an inverse correlation was found between sputum neutrophils and *Moraxella catarrhalis*. Our results also support the potential association between *Moraxella catarrhalis* and neutrophilic inflammation induction in severe asthmatics; when looking at the molecular phenotype level, the abundance of *Moraxella catarrhalis* was highest in the TAC2 phenotype, which is characterized by activation of neutrophils and the inflammasome. This finding, in combination with its correlation with sputum neutrophilia, further supports the link between *Moraxella catarrhalis* and neutrophilic inflammation. In the same vein, *Moraxella catarrhalis* has been shown to act on several mediators of neutrophilic inflammatory processes, such as macrophage activation through TLR 2, 4 and 9 (231) as well as NF-κB-dependent release of IL- 6, IL- 8 and prostaglandin E2 in the lung epithelium (232,233).

While commonly associated with Whipple’s disease with inflammatory involvement of the gastrointestinal system, *Tropheryma whipplei* has also been implicated in the aetiology of several respiratory conditions; aspiration as well as pneumonia (both ventilator-associated and community-acquired), can be caused by *Tropheryma*

whipplei (234). When examining bronchoalveolar lavage fluid of asymptomatic immunosuppressed Human Immunodeficiency Virus patients, *Tropheryma whipplei* colonization was evident and was susceptible to anti-retroviral therapy (235). There is also a report of *Tropheryma whipplei* in sputum samples of severe asthmatic with a predominately eosinophilic phenotype, as determined via PCR (134). Our findings revealed that among severe asthma patients, current smokers or ex-smokers have the highest levels of *Tropheryma whipplei*. Patients from the Sas/ex are more likely to exhibit sputum eosinophilia when higher abundance of *Tropheryma whipplei* was present. Moreover, pack years of smoking also affected the abundance of *Tropheryma whipplei* in a positive trend among patients presenting with the TAC1 and TAC2 phenotypes. At the time of writing, this remains the only evidence of the implication of smoking status on sputum metagenome of severe asthmatics (particularly, Sas/ex vs non-smoking asthmatics). Previous examinations of the lung microbiome of smokers in comparison with non-smokers employing 16S ribosomal RNA microarray have either failed to detect a difference between the two groups (236,237), or have reported on the level of microbiota other than *Tropheryma whipplei* (increase in *Veillonella* and *Megasphaera* with a decrease in *Haemophilus*) (237). As such, it is uncertain what mechanisms underly the increased abundance and potential pathogenesis of *Tropheryma whipplei* in smoking/ex-smoking severe eosinophilic asthma.

The above-described variations in bacterial species were not the only ones found in the analysis; Severe asthmatics, be they smokers (Sas/ex group), ex-smokers (Sas/ex group), or non-smokers (SAn group), exhibited reductions in several species when compared to the MMA and/or HC groups. Consistently with extant evidence of enriched *Prevotella* in non-asthmatics compared to asthmatics (56,138), the present analysis noted a decreased abundance of *Prevotella intermedia* in patients with severe asthma, regardless of their smoking status (SAn and SAS/ex groups). Moreover, a negative correlation was observed between most of these bacterial species and sputum neutrophilia. Similar associations emerged on the level of asthma exacerbations; an increase in the abundance of *Prevotella melaninogenica* and *Prevotella intermedia* was indicative of lower asthma exacerbations in the whole asthma group. As such, samples from patients with non-smoking severe asthma had

reductions in some *Prevotella* species (vs. HC and MMA), consistent with previous reports of higher levels in healthy controls and mild-moderate asthmatics (56). That being said, knowledge on these bacterial species' potential lung pathogenicity is limited. In particular, few studies have examined the implication of *Prevotella* species in the human respiratory system, although it has been linked to mucosal inflammation through Th-17 mediated pathways (238).

Another finding of note is the notable reduction of *Haemophilus parainfluenzae* in non-smokers (SAn) and in both TAC1 and TAC2. As for this species' relationship with sputum neutrophilia, the SAn and MMA groups, as well as the TAC1 and TAC3 had negative correlations. While current smokers also had reduced abundance of *Haemophilus parainfluenzae* when compared to ex-smokers, patients on OCS exhibited similarly reduced levels compared to those who are not, regardless of whether they were current or ex-smokers. This calls into question the role of current smoking in reducing this bacterial species as this particular change in the bacterial microbiome could be due to OCS therapy. It is worth mentioning that by triggering TAK1/MAPK activation, *Haemophilus parainfluenzae* can negatively affect corticosteroid responsiveness in asthmatic patients' alveolar macrophages (110), possibly contributing to the development of corticosteroid resistance.

Follow-up data after 12 months suggests the stability of the microbiome in severe asthma patients, as determined by the maintenance of α- and β-diversity. Relative overtime stability was evident in two distinct robust severe asthma phenotypes when the same microbial data was subjected to unbiased microbiome-driven clustering (87). However, achieving relative stability does not exempt asthma patients from experiencing worsening of their disease, such as acute exacerbations, and consequently exhibit microbiome changes (specifically, an increase in gram-negative microbes) in induced sputum (239). Delving deeper into the two identified phenotypes, it was found that 40 patients exhibited worse asthma outcomes, increased sputum neutrophilia and greater abundance of several species (i.e. gamma-proteobacteria, *Haemophilus influenzae* and *Moraxella catarrhalis*). A genetic component has been linked to these phenotypes, with previous report of changes in the expression of genes implicated in inflammatory processes and their regulation such as TNFα, IL-1, TLRs and inflammasomes (240).

Some technical limitations of the analysis presented above are worth mentioning. Firstly, MetaPhlAn2 was used based on the higher utilitarian value of its taxonomic composition for respiratory microbiome consideration compared to protein-based methods, such as Kaiju (211). Despite this, higher accuracy in taxonomy profiling could have been achieved with sequencing technology reads of 150 base-pairs or longer at lower depth compared to the paired-end 100 base-pairs at high sampling depth used in our analysis.

Secondly, the taxonomic assignment approach used in MetaPhlAn2 (algorithmic marker-based approach) may not yield complete results; This is thought to arise out of the generation of high amounts of host DNA in sputum samples by neutrophils and neutrophil extracellular traps and the consequent low level of microbial reads. The estimation of taxonomic composition thus becomes difficult based on the resulting lack of marker genes, leading the exclusion of some samples from the downstream analysis. DESeq2 was used for differential analysis in order to account for the negative binomial distribution of metagenomic data as well as make allowances for pseudocount, thus addressing the fact that not all samples share the same taxa. Another important limitation in our study was incomplete data on previous or current medication (antibiotics, oral corticosteroid therapy, etc.) which could impact lung microbiota. Selective reduction of bacterial load has been reported with some therapy, with a study reporting azithromycin (macrolide antibiotic) to lead to a significant reduction in *Haemophilus Influenzae* among patients with unstable asthma. However, no decrease was observed in *Streptococcus pneumoniae, Staphylococcus aureus, Pseudomonas aeruginosa,* or *Moraxella catarrhalis* (241). In the current patient cohort, few reported antibiotic therapy at baseline or follow-up, but treatment history was unavailable. Moreover, results indicated a decrease in *Haemophilus Influenzae* and *Moraxella catarrhalis* with the use of OCS in severe asthmatics who were non-smokers, further supporting the role of OCS, and other relevant therapy, in the modulation of the microbiome.

In conclusion, this metagenomic analysis of sputum derived from severe asthmatics illustrated the airway inflammation phenotype-dependent change in bacterial α-diversity at the species level. On the one hand, neutrophilic inflammation and activation of neutrophils and inflammasome were associated with reduced

abundance of some bacterial species, namely *Haemophilus influenzae* and *Moraxella catarrhalis*. On the other hand, *Haemophilus influenzae* and *Tropheryma whipplei* were most abundant in the presence of eosinophilic inflammation, while *Moraxella catarrhalis* could not be detected. Moreover, associations emerged between *Tropheryma whipplei* and eosinophilic inflammation, as well as smoking (be it current or ex-smoking). The associations revealed in our analysis suggest the potential causal relationship between specific inflammatory pathways and bacterial species; particular inflammatory pathways could be active by some bacteria, while the opposite holds true with certain inflammatory pathways inducing selective bacterial proliferation. Regardless, results suggest the correlation between disease severity and inflammatory phenotype (especially neutrophilic inflammation) and bacterial species established in the airways of severe asthmatics. Future studies are needed in order to confirm and explore the implication of various bacterial species in the inflammatory processes underlying asthma, particularly by evaluating variations on the lung microbiome between patients treated with type 2 or non-type 2 inflammation pathway treatments. There is current evidence to suggest that certain clinical variables such as fractional exhaled nitric oxide (FeNO) might be able to predict which subgroup of severe asthma is more likely to benefit from antimicrobial strategies, and that biologic therapy used to improve asthma control in type 2-mediated severe asthma, such as mepolizumab, does not necessarily alter airway microbial composition (113). Further research is needed to better understand the optimal integration of microbial species (and their targeted treatments) in asthma patients' management.

Chapter 5. A severe neutrophilic asthma phenotype with high airway Haemophilus influenzae abundance

Parts of this chapter are based on the publication:

***A Versi**, A Azim et al. A severe asthma phenotype of excessive airway Haemophilus influenzae relative abundance associated with sputum neutrophilia. Clin Transl Med. 2024 Sep;14(9):e70007.*

5.1 Introduction

A paradigm shift in asthma unfolded with the emergence of molecular phenotypes in asthma, characterized by the predominant type of inflammation (eosinophilic or neutrophilic inflammation) that are established based on transcriptomic or proteomic analyses of severe asthmatics' bronchial biopsies or sputum cells (20,175). In fact, molecular phenotypes carried important implications for the management of severe eosinophilic asthma; this phenotype is characterized by type 2 inflammation and thus exhibited a good response to biologic therapy targeting type 2 inflammatory mediators such as cytokines, IL-4, IL-5 and IL-13. By contrast, studies have yet to clearly define neutrophil-driven asthma phenotypes beyond their potential association with airway microbiome dysbiosis which is common to uncontrolled or severe asthma (161,240). Species-level metagenomic insights into the airway microbiome of asthma patients has linked neutrophilic asthma to an abundance of *Haemophilus influenzae* and *Moraxella catarrhalis (170)*, consistently with other reports of *Proteobacteria* enrichment (including *Moraxella* and *Haemophilus*, particularly *Haemophilus influenzae*) in the neutrophilic asthma phenotype (134,161). It seems that *Haemophilus influenzae* and *Moraxella catarrhalis* are involved in triggering an inflammatory response via the activation of neutrophils and inflammasomes (170). That being said, opportunistic pathogens have been known to develop evasive mechanisms for the evasion of neutrophil-driven bacterial killing, thus circumventing the crucial role of neutrophils in the defence against respiratory infections (242).

Based on the evident interplay between pathogenic bacteria and the host immune response in the airways, we examined whether the abundance of bacterial species

could further the definition of severe asthma phenotypes. To that end, a definition of bacterial species outgrowths was developed based on the impairment of the diversity of species revealed by the metagenomic analysis of sputum samples in the severe asthma U-BIOPRED cohort (170). We assessed the associations between host molecular pathways and relative dominant species containing specific abundant bacterial species. In so doing, we have defined a phenotype of severe asthma dominated by abundant airway *Haemophilus influenzae*.

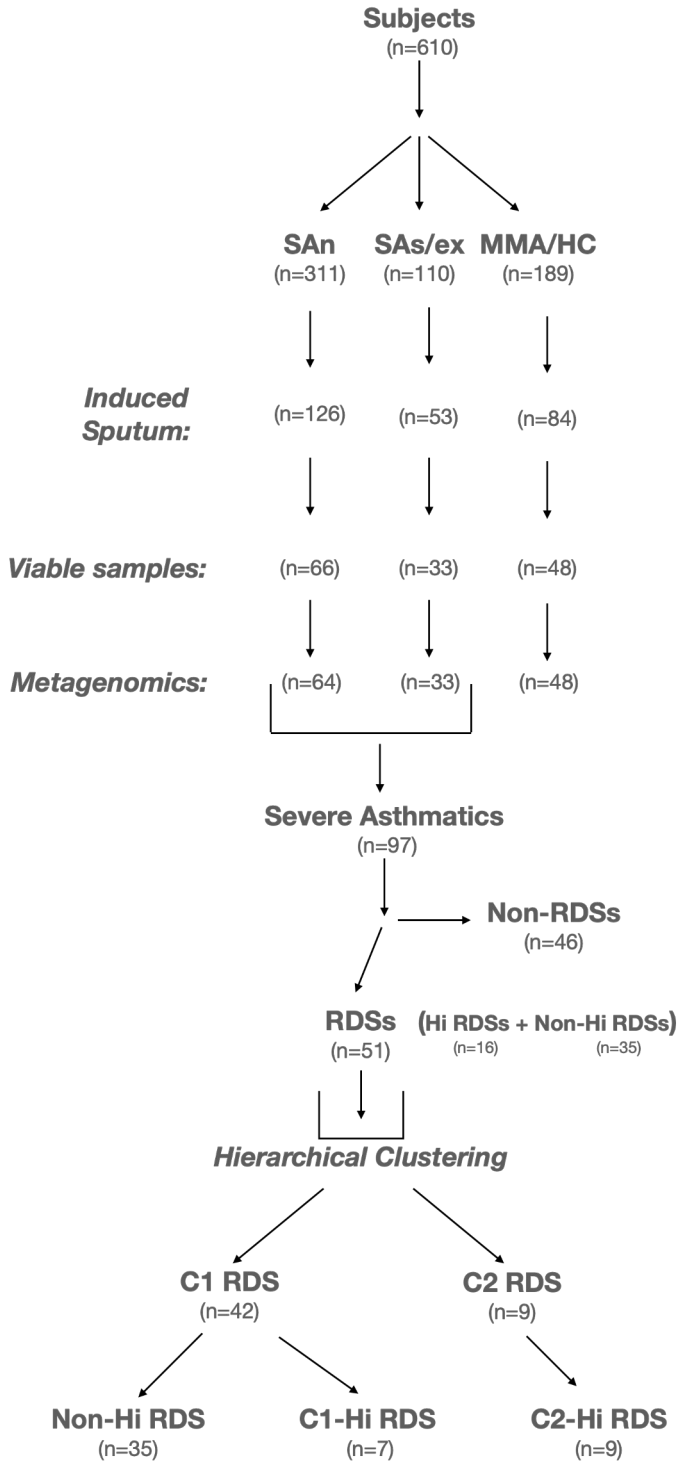
The hypothesis of this chapter was that an increased relative abundance of particular species in the airway microbiome may be associated with distinct host immune response pathways.

5.2 Method

5.2.1 Participants

The U-BIOPRED cohort included severe asthmatics (SA) and two control groups. SAs consisted of nonsmokers and current and/or ex-smokers, while mild-moderate asthmatics (MMA) and healthy volunteers (HC) constituted the two control groups (164) (**Table 6**). Written informed consent was obtained from all patients prior to study participation.

Figure 27. Consort diagram depicting subjects in analysis



Legend: Consort diagram depicting the repartition of asthma subjects recruited in U-BIOPRED cohort into various groups and in the hierarchical clustering with 3 clusters of RDS.

Hi: *Haemophilus influenzae*; MMA/HC: mild-moderate asthma and Healthy control subjects; RDS: relative dominant species; SAn: severe non-smoking asthma; SAs/ex: severe ex-smoking or current smoking asthma.

5.2.2 Sample collection, DNA/RNA extraction and data pre-processing

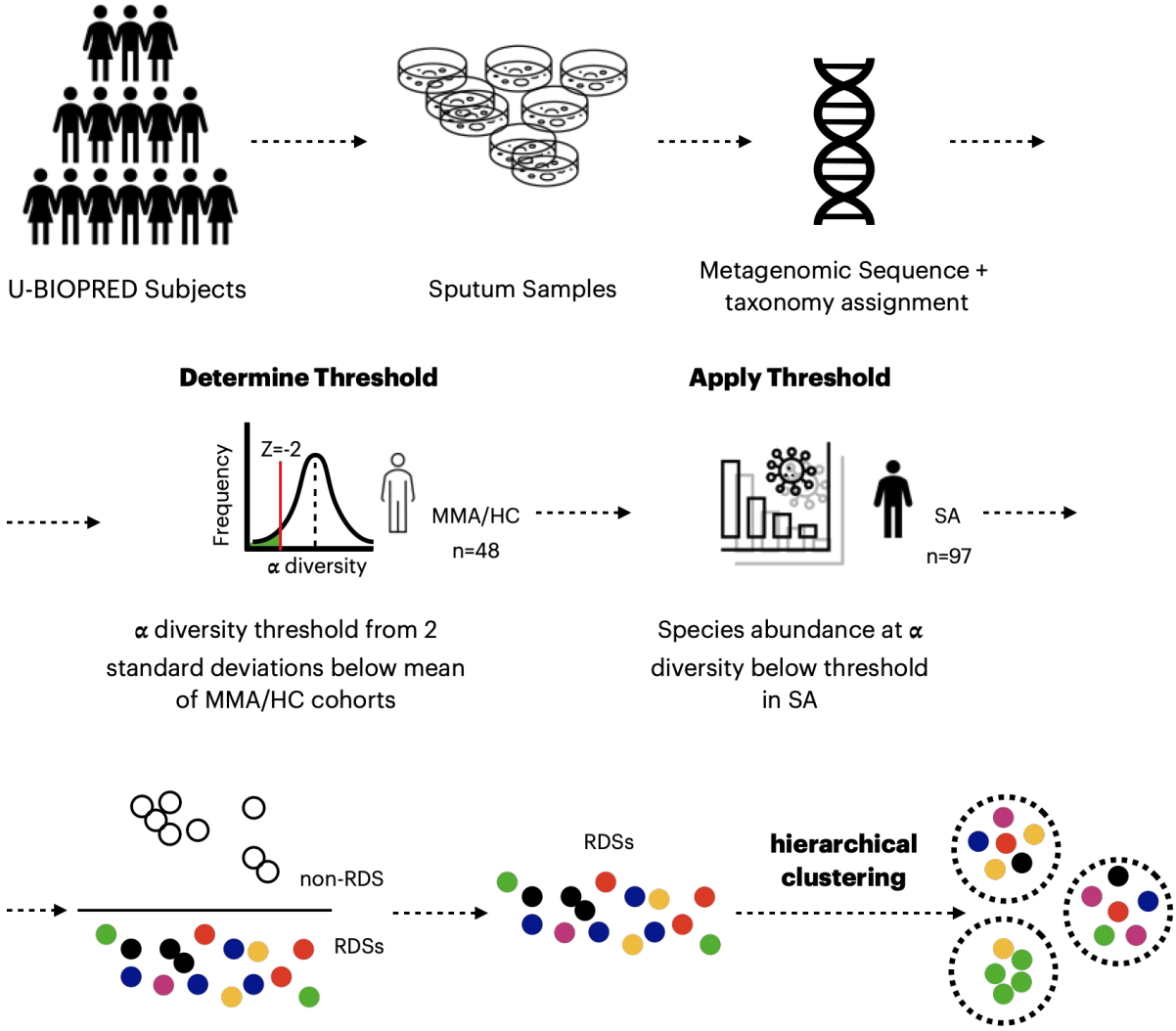
The patients inhaled nebulized hypertonic saline for the induction of sputum. Sputum cells from the collected induced samples were used for RNA extraction. Transcriptomic expression was performed using this RNA via Affymetrix U133 Plus 2.0 (Affymetrix, Santa Clara, CA, USA) microarrays. Quality checks were performed according to Affymetrix® recommendations and the expression matrix was derived using the robust microarray analysis (RMA) method from the *affy R package (243)*.

Metagenomic sequencing analysis was performed on frozen sputum samples from 148 participants (100 SA, 25 MMA, 23 HC) by Second Genome (San Francisco, California, USA) (170) (**Figure 27**). Following this, metagenomic data was subjected to quality control and removal of host reads yielding a final sample of 145/148 (97 SA patients and 48 MMA and HC subjects) for further analyses. Microbiome profiles were then estimated through the MetaPhlAn2 pipeline (*version 2.7.2*) and its marker database (220). Metagenomic functional profiling was undertaken to estimate the abundance of microbial gene families and pathways from metagenomic data using the HUMAnN2 (213) and the UniProt pathway database (UNIPATHWAY database) (214,244).

5.2.3 Definition of relative dominant species (RDSs)

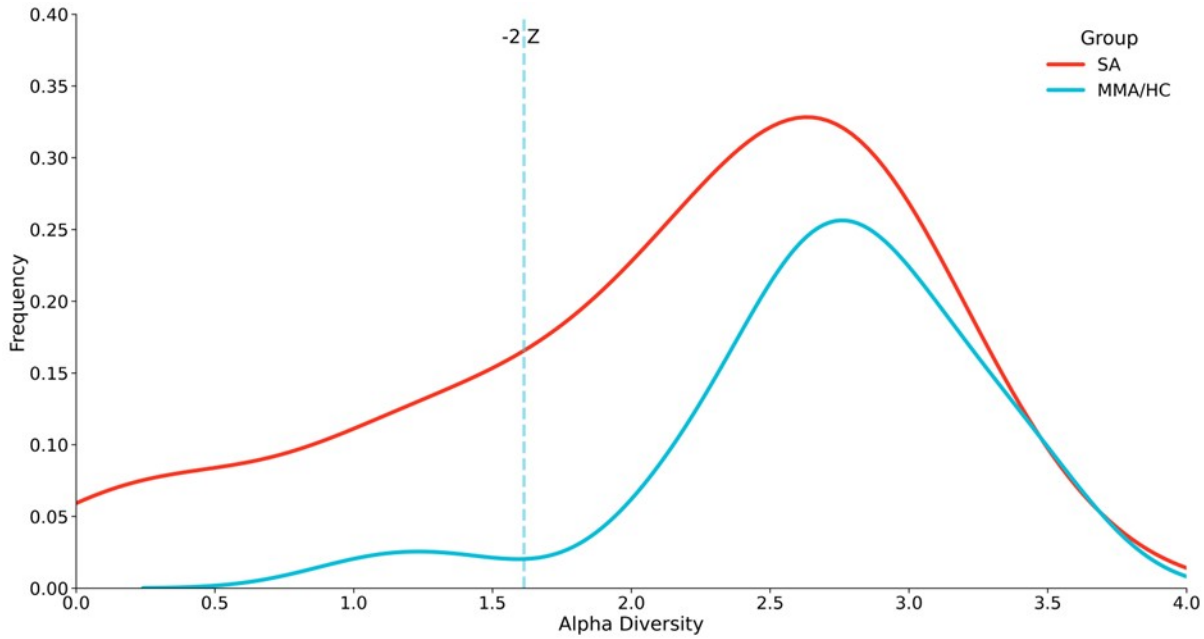
We defined the abundance of a particular bacterial species being high to be designated as an RDS. In the SA cohort, for each species a minimum abundance cut-off was determined based on the minimum abundance required to for a reduction in Shannon's α-diversity to below the Shannon's α-diversity pre-determined threshold of a Z-Score of -2 obtained from the combined MMA and HC cohorts (**Figure 28** & **Figure 29**). Samples with an abundance of equal or above the mimimum abundance cut-off for that species would be designated as an RDS.

Figure 28. Flowchart showing the determination of relative dominant species (RDS) by α-diversity threshold in the sputum samples and applied to samples of severe asthma patients and hierarchical clustering of RDS samples.



Legend: MMA/HC: Mild-moderate asthma/Healthy controls; SA: Severe asthma. The numbers studied in each cohort are shown.

Figure 29. Relative dominant specie (RDS) cut-off methodology.



Legend: The α-diversity (Shannon's diversity index) distribution of mild-moderate asthma/healthy controls (MMA/HC) and severe asthma (SA). Those with α-diversity in the range of or less than –2-Z score were included as RDSs.

5.2.1 Method for designating a relative dominant species (RDS)

The formula for the Shannon's diversity index is described below:

where p_i denotes the proportion of the entire sample composed of species *I (245).*

Used to measure entropy, the shannon's diversity index originates in information theory. The formula above illustrates the concept that as the number of taxa or species there are and as their equalness in abundance increases, then it becomes more difficult to predict the next taxa/species and its abundance - hence there is higher entropy or disorder in the sample (246). Therefore, the Shannon's α diversity in the current context is a metric of richness and evenness of the species in a

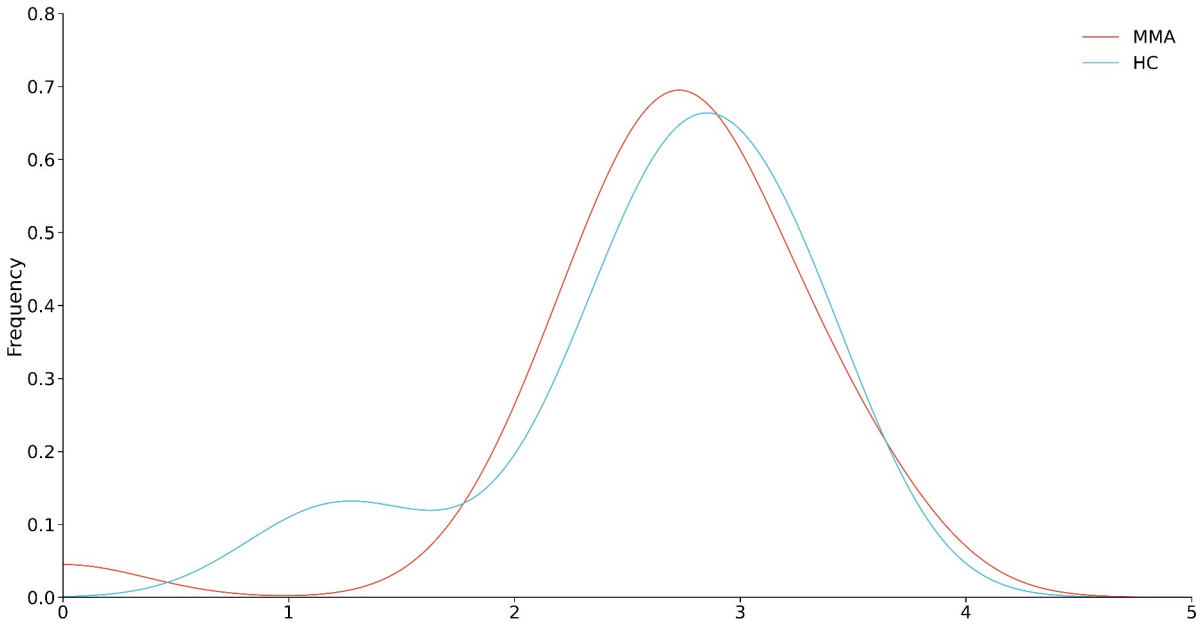
particular sample. The term $-p_i ln(p_i)$ contributes the measure of evenness in the equation.

As decrease in Shannon's diversity denotes, a decrease in evenness and richness of species within the sample. A decrease in evenness of abundance within a sample denotes that the abundance of one or more species rises above the rest of the sample, hence a dominant species emerges. To determine an arbitrary point at which the evenness is reduced enough to be abnormal, a 2-standard deviation below the mean of the mild-moderate asthma and healthy control (MMA/HC) group was used as an arbitrary threshold. By using this threshold on the severe asthma group we were able to identify the species which are attributed to this low evenness within the samples. Those samples within severe asthma that do not fall below this threshold would be considered not to have severe asthma associated with any degree of microbial dysbiosis.

This approach enables us to identify the bacterial species which are associated with this reduction in evenness within the sample and can then be defined as a relative dominant species (RDS). These samples with a particular RDS can then be analysed with paired transcriptomics data to gain insight into the host response to the bacterial species dominance. This makes it possible to explore the heterogeneity of the host response to a particular species.

As the aim of this study was to gain a deeper understanding of the abnormalities associated with low diversity severe asthma, the healthy control (HC) subjects and the mild-moderate asthma (MMA) subjects were combined into a single comparator group. This approach was deemed appropriate because the mean α diversity of MMA and HC were within 5% of each other with the standard deviation of MMA cohort less than the standard deviation of the HC cohort. The Z-test between the two distributions of MMA vs HC showed they were not significantly different, with a Z-statistic of -0.3, suggesting the samples were of the same distribution and a non-significant p-value (0.76). **Figure 30** depicts the distribution plots of the two cohorts.

Figure 30. Distribution plot of alpha-diversity for MMA and HC Cohorts



Legend: A Frequency plot of the alpha diversity using Shannon's index for the cohorts MMA and HC. MMA: Mild-moderate asthma; HC: Healthy controls.

Using this diversity threshold in the SA group, an abundance cut-off was determined for each bacterial species based on the minimum abundance of that species associated with the reduction in Shannon's a-diversity to, or below, this threshold (Method for designating species RDS). We iterated through a set of abundance values for each species and excluded samples that do not have an abundance at or above that value and then calculated the diversity of the remaining samples. When the Shannon's α-diversity of the remaining samples was found to be at or below the pre-calculated threshold, that abundance value was designated as the RDS abundance cut-off for that species. For samples that contained no RDSs, the species was designated as a non-RDS. The algoritm is shown below using pseudo-code.

ALGORITHM 1. ALGORITHM FOR RDS DETERMINATION
Input**: abundance_matrix, Shannon_threshold* ***N <- input.length ***For species = 0 to N:***

```
        For cutoff = 0 to 75 with step 0.5:
            new_abundance_matrix <- select_samples_of _species_less_than(species, cutoff)
            If mean_shannon_diversity(new_abundance_matrix) >= shannon_threshold:
                Output <- [species, cutoff]
                break
        End
    End
```

5.2.2 Clustering on abundance of low-diversity bacterial species

Clustering by similar species compositional characteristics was performed. To determine similarity between samples, the Aitcheson distance β-diversity dissimilarity measure (247) was employed because it was suited for compositional data (248). This was computed on the species-level abundance from the metagenomics data of the RDS samples in SA. Hierarchical Ward2 agglomerative clustering on the Aitcheson distance was performed and the optimum number of clusters determined using the average silhouette width (249) and the Calinski-Harabasz score (250). Consensus clustering (15) was used to assess stability by resampling, randomly removing 10% of the data and repeating the clustering through 1,000 iterations. The empirical CDF displays the consensus distributions for each value of k derived using the ConsensusClusterPlus package (251).

5.2.1 Determiniation of dominant RDS

For samples containing multiple RDSs, the definition of a dominant RDS is described. We divided the abundance of all RDS species within a sample by the cut-off abundance for that given species. This produced a factor and the spcies with the

highest factor in a given sample was considered to be the dominant RDS in that sample.

5.2.2 Longitudinal analysis

Paired baseline samples were available for 43 out of the 44 follow-up samples. This was used to identify the stability of dominant RDSs between baseline and follow-up.

5.2.3 Statistical analysis

We used *pandas 1.4.2* (180) and *python 3.9.10* (181) data analysis tools. Statistical analyses were performed using R version 4.1.1 (179). Differentially-abundant microbiome taxa (DAT) between clusters were identified using ANCOM-BC (252) which was selected due to its suitability for compositional data (248). The Holm-Bonferroni for adjustment was used to control for false discovery rate (FDR). DAT analysis was performed between clusters generated from the hierarchical clustering of low-diversity RDS samples and MMA/HC group.

DEG analysis was performed on sputum transcriptomics (253) between different RDS species vs MMA/HC, and non-RDS vs MMA/HC groups using Limma (183). A Benjamini–Hochberg FDR adjustment was applied with FDR <0.05 and $\log_2$ fold change ≥ 0.5 were considered statistically significant in transcriptomic analyses. DEG results were used to perform GSEA using ClusterProfiler (254). Comparison of the clinical variables between groups was performed using analysis of variance (ANOVA) for multiple group comparison of normally-distributed variables. Kruskal-Wallis test was used for multiple group comparison of non-normally distributed variables or ordered categorical and chi-squared test was used for qualitative variables.

5.3 Results

5.3.1 Low a-diversity RDSs

Sputum neutrophil levels and sputum macrophages were lower in patients with normal a-diversity compared to those with low a-diversity, who were more likely to have allergic rhinitis **(Table 6**).

Table 6. Characteristics of relative dominant species (RDS) and non-RDSs.

	Baseline			Longitudinal		
	RDSs	Non-RDSs	P Value	RDSs	Non-RDSs	P Value
Subjects, n	51	46	NA	24	19	
Age (years)	54.0 [47.0,59.5]	55.0 [45.2,62.8]	NS	56.0 [46.8,61.2]	58.0 [53.5,65.5]	NS
Female, n (%)	27 (52.9)	29 (63.0)	NS	12 (50.0)	12 (63.2)	NS
BMI	28.6 (5.1)	28.8 (6.3)	NS	27.8 (4.9)	28.1 (5.1)	NS
Duration years,	27.9 (16.1)	22.4 (16.1)	NS	28.9 (15.2)	21.5 (17.4)	NS
Current smoking, n (%)	5 (9.8)	6 (13.0)	NS	2 (8.3)	1 (5.3)	NS
Oral corticosteroid daily, n (%)	22 (43.1)	17 (37.0)	NS	8 (33.3)	4 (21.1)	NS
ACQ-5 score,	2.3 (1.2)	2.3 (1.2)	NS	2.1 (1.3)	1.8 (1.2)	NS
Exacerbations, per year,	2.1 (2.2)	2.4 (2.1)	NS	1.5 (1.8)	2.3 (2.3)	NS

Nasal polyposis, n (%)	18 (35.3)	15 (32.6)	NS	7 (29.2)	9 (47.4)	NS
Eczema, n (%)	14 (27.5)	11 (23.9)	NS	7 (29.2)	3 (15.8)	NS
Allergic rhinitis, n (%)	25 (49.0)	11 (23.9)	<0.05	10 (40.0)	6 (31.6)	NS
GERD, n (%)	24 (47.1)	18 (39.1)	NS	12 (48.0)	8 (42.1)	NS
Blood Eosinophils μL^{-1}	0.3 (0.3)	0.3 (0.3)	NS	0.2 (0.1)	0.4 (0.2)	<0.005
Blood Neutrophils μL^{-1}	5.7 (2.5)	5.1 (2.6)	NS	6.0 (2.8)	4.7 (1.9)	NS
Sputum Neutrophils (%)	63.6 (24.7)	49.5 (24.0)	0.005	65.8 (20.5)	55.1 (23.7)	NS
Sputum Eosinophils (%)	14.5 (20.2)	13.1 (20.5)	NS	8.1 (13.6)	14.9 (22.2)	NS
Sputum Lymphocytes (%)	1.0 (1.1)	1.6 (1.4)	<0.05	1.3 (1.7)	1.9 (1.5)	NS
Sputum Macrophage (%)	20.8 (18.5)	35.7 (22.5)	0.001	24.5 (20.7)	28.1 (17.2)	NS
FEV1% predicted	62.4 (21.3)	67.4 (18.7)	NS	59.9 (17.5)	64.9 (15.6)	NS
FeNO (ppb),	35.0 (34.9)	40.1 (38.6)	NS	31.5 (26.9)	50.9 (46.0)	NS
Smoking (pack	15.3	18.3	NS	17.4	12.5 (6.6)	NS

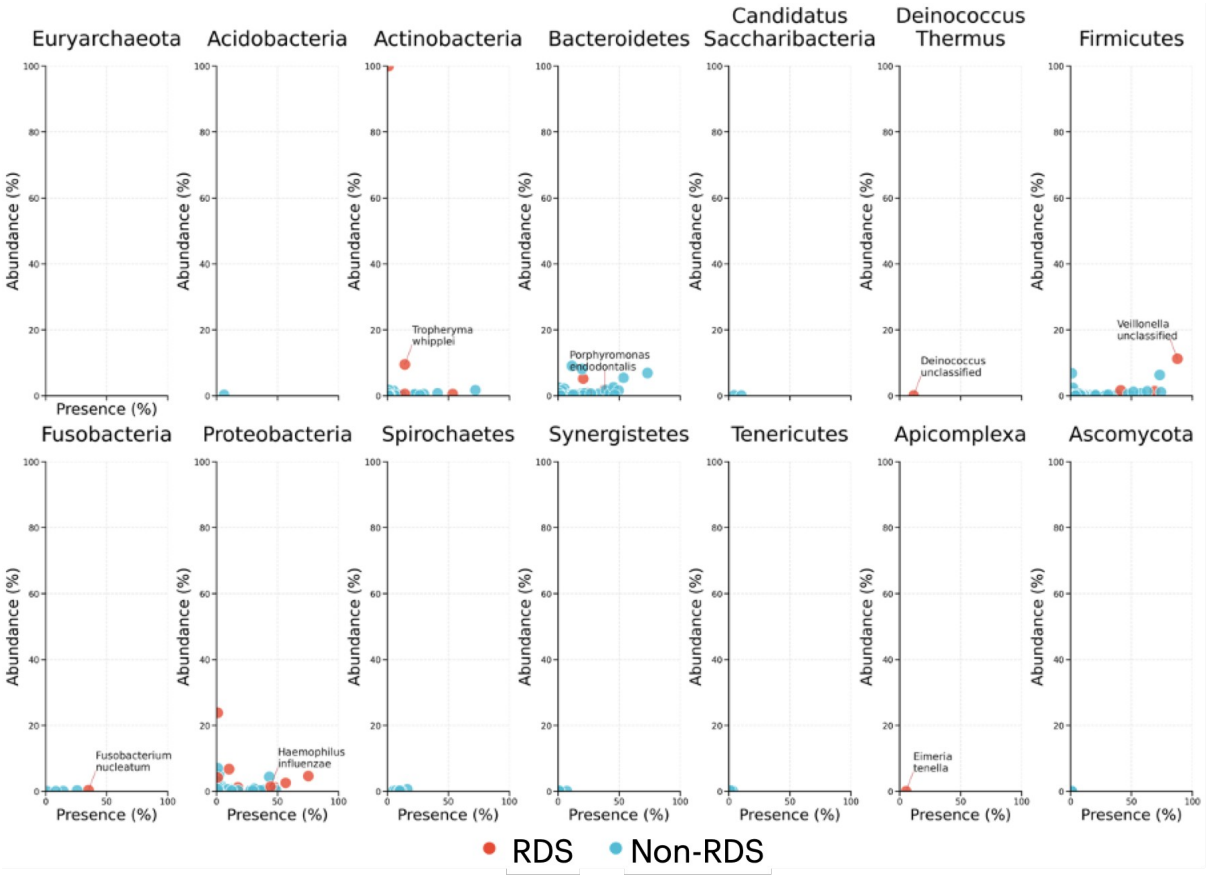
years)	(16.1)	(18.4)		(19.3)		
Antibiotic current, n (%)	7 (13.7)	10 (21.7)	NS	4 (16.0)	3 (15.8)	NS
Serum CRP (mg/L)	6.6 (12.2)	6.6 (15.3)	NS	6.4 (11.2)	6.8 (19.5)	NS
Serum IL1α (pg/ml)	35.7 (7.9)	35.1 (6.4)	NS	36.5 (9.4)	35.2 (7.1)	NS
Serum IL6 (pg/ml)	1.4 (1.2)	1.1 (0.8)	NS	1.0 (0.7)	1.2 (0.7)	NS
Serum IL8 (pg/ml)	4.8 (7.0)	4.0 (3.0)	NS	6.0 (10.2)	4.9 (4.3)	NS
Serum C5a (pg/ml)	50.4 (28.8)	39.8 (19.6)	<0.05	54.8 (32.8)	38.9 (20.5)	NS

Abbreviations:

ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; CRP: Serum C-Reactive Protein; FeNO: Fractional exhaled nitric oxide; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; GERD: Gastroesophageal reflux disease; N/A: Not applicable; NS: Not significant; ppb: parts per billion. Data shown as mean (Standard Deviation) unless variable is categorical where n (%). Age shown as median [95% Confidence intervals].

P value: One-way ANOVA (continuous) & Chi-squared (categorical) tests p-value comparing RDSs and non-RDSs.

Figure 31. Presence and abundance of bacterial species by phylum.



Legend: Scatter plot showing the presence and abundance of all species classified by the species phylum in the severe asthma groups, SAn and SAs/ex. Presence is defined as the percentage of samples where a particular species exists with a greater than zero abundance. Abundance is defined as the median abundance of all the samples for a particular species where it exists with a greater than zero abundance. Red dots indicate species relative dominant species, and blue indicates species without relative dominant species.

Overall, 24 species were categorised as being RDSs. The most common RDSs in descending order were *Haemophilus influenzae* (16 samples), *Actinobacillus unclassified* (n=10), *Veillonella unclassified* (n=9), *Haemophilus aegyptius* (n=9), *Streptococcus pseudopneumoniae (n=7), Propionibacterium acnes (n=5), Moraxella catarrhalis (n=5)* and *Tropheryma whipplei* (n=5) (

).

Table 7. Presence and abundance of bacterial species in sputum of severe asthma.

Species	Phylum	RDS Samples (n=97)	Presence (% of samples with detectable species)	Mean abundance (%)*	Mean abundance in RDS samples (%)*
Haemophilus influenzae	Proteobacteria	16	44.33 (43/97)	18.40 ± 31.73	48.23 ± 36.14
Actinobacillus unclassified	Proteobacteria	10	47.42 (46/97)	2.00 ± 2.73	5.62 ± 4.05
Veillonella unclassified	Firmicutes	9	87.63 (85/97)	14.00 ± 17.10	50.53 ± 30.96
Haemophilus aegyptius	Proteobacteria	9	17.53 (17/97)	2.02 ± 2.07	3.40 ± 1.98
Streptococcus pseudopneumoniae	Firmicutes	7	41.24 (40/97)	8.48 ± 17.04	40.05 ± 21.40
Moraxella catarrhalis	Proteobacteria	5	10.31 (10/97)	24.98 ± 37.50	49.09 ± 41.35
Tropheryma whipplei	Actinobacteria	5	14.43 (14/97)	22.13 ± 25.67	53.71 ± 12.31

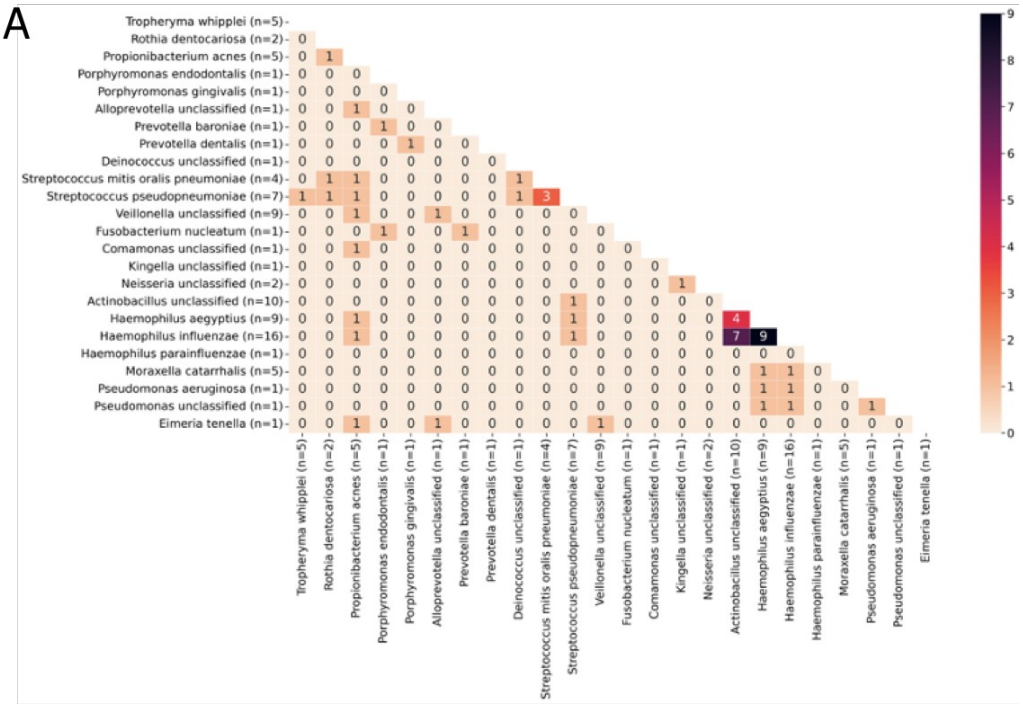
Propionibacterium acnes	Actinobacteria	5	14.43 (14/97)	7.24 ± 16.22	19.77 ± 23.45
Streptococcus mitis oralis pneumoniae	Firmicutes	4	69.07 (67/97)	3.80 ± 6.82	26.10 ± 5.71
Neisseria unclassified	Proteobacteria	2	56.70 (55/97)	5.46 ± 8.22	38.98 ± 4.06
Rothia dentocariosa	Actinobacteria	2	53.61 (52/97)	2.11 ± 5.65	28.79 ± 6.08
Pseudomonas aeruginosa	Proteobacteria	1	1.03 (1/97)	23.83	23.83
Porphyromonas gingivalis	Bacteroidetes	1	20.62 (20/97)	13.04 ± 18.85	75.01
Haemophilus parainfluenzae	Proteobacteria	1	75.26 (73/97)	6.18 ± 6.80	46.75
Porphyromonas endodontalis	Bacteroidetes	1	38.14 (37/97)	5.48 ± 12.25	66.76
Pseudomonas unclassified	Proteobacteria	1	1.03 (1/97)	4.23	4.23
Alloprevotella unclassified	Bacteroidetes	1	23.71 (23/97)	3.02 ± 6.71	32.32
Deinococcus unclassified	Deinococcus Thermus	1	11.34 (11/97)	1.72 ± 5.26	17.57
Kingella unclassified	Proteobacteria	1	34.02 (33/97)	1.63 ± 5.32	30.86
Prevotella baroniae	Bacteroidetes	1	15.46	1.53 ±	12.70

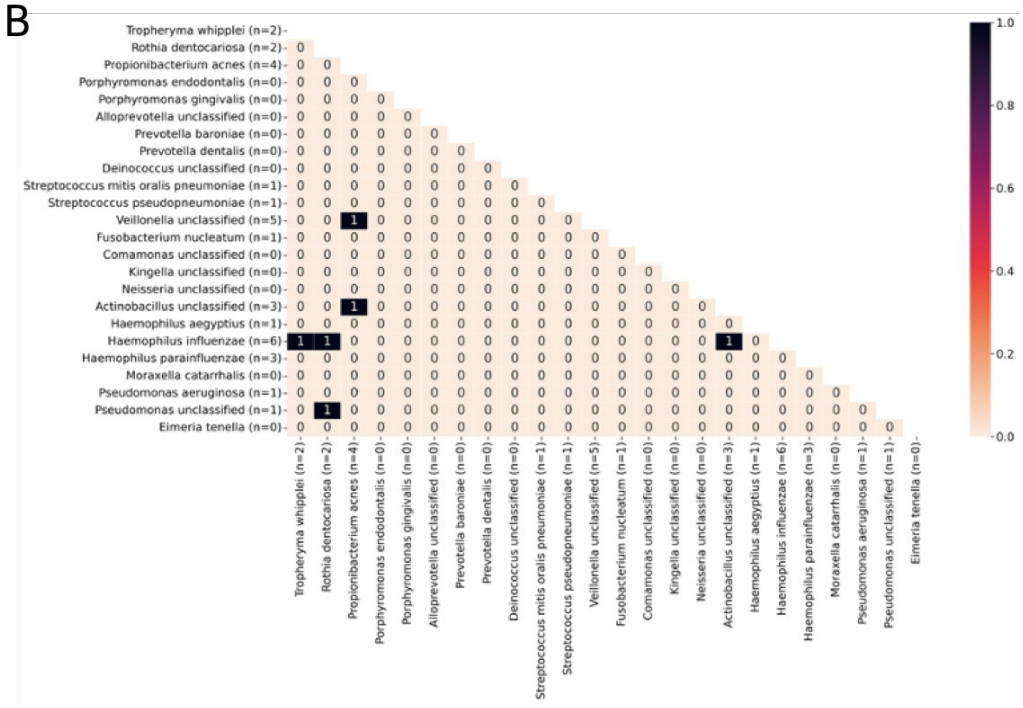
			(15/97)	3.49	
Fusobacterium nucleatum	Fusobacteria	1	35.05 (34/97)	1.03 ± 2.15	12.16
Comamonas unclassified	Proteobacteria	1	3.09 (3/97)	0.53 ± 0.90	1.57
Prevotella dentalis	Bacteroidetes	1	14.43 (14/97)	0.50 ± 0.80	3.11
Eimeria tenella	Apicomplexa	1	5.15 (5/97)	0.26 ± 0.50	1.15

Legend: Presence denotes the percentage of samples where a particular species exists with a greater than zero abundance. Abundance is the median abundance of all the samples for a particular species where it exists with a greater than zero abundance.
RDS: Relative dominant specie. * Data shown as mean ± SD.

As shown in **Table 6** close to half (52.6%) (51 out of 91) of samples contained at least one RDS. That being said, the RDS were in the majority mutually exclusive; of the 97 samples, 29 contained only one RDS, 10 samples two RDSs, 9 samples three RDSs and 3 samples four RDSs (**Figure 32A).** *Haemophilus aegyptius* RDSs always co-occurred with *Haemophilus influenzae* RDSs or the *Actinobacillus unclassified* RDS. The *Actinobacillus unclassified* RDS generally co-occurred with the *Haemophilus influenzae* RDS *(7 samples). Pseudomonas aeruginosa and Pseudomonas unclassified RDSs* were seen in 1 sample and co-occurred with the *Haemophilus aegyptius* RDS and *Haemophilus influenzae* RDS*. Moraxella catarrhalis* RDS only co-occurred with one or more RDS species in 1 sample, which were the RDS species *Haemophilus aegyptius* and *Haemophilus influenzae. Tropheryma whipplei* RDS cooccurred with the *Streptococcus pseudopneumoniae* RDS in 1 sample.

Figure 32. Relative dominant specie (RDS) co-occurrence heatmap.





Legend: The heatmap shows the co-occurrences of RDSs within severe asthma groups (combined SAn and SAs/ex). The heatmap defines the number of samples containing more than one RDS type. The number of samples containing a particular RDS is denoted in the labels of the plot for each RDS. **Panel A** is for baseline and **Panel B** is for follow-up.

5.3.2 Haemophilus influenzae, Moraxella catarrhalis and Tropheryma whipplei RDSs

5.3.2.1 *Haemophilus influenzae RDS*

Haemophilus influenzae was detectable in 44% (43/97) of severe asthmatic (SAn & SAs/ex) samples with a mean abundance of 1.42% ± 13.5%. 37% (16/43) of samples with *Haemophilus influenzae* met the loss of diversity criteria, becoming a RDS. The mean abundance of *Haemophilus influenzae* within RDS samples was 47.02 ± 36.14.

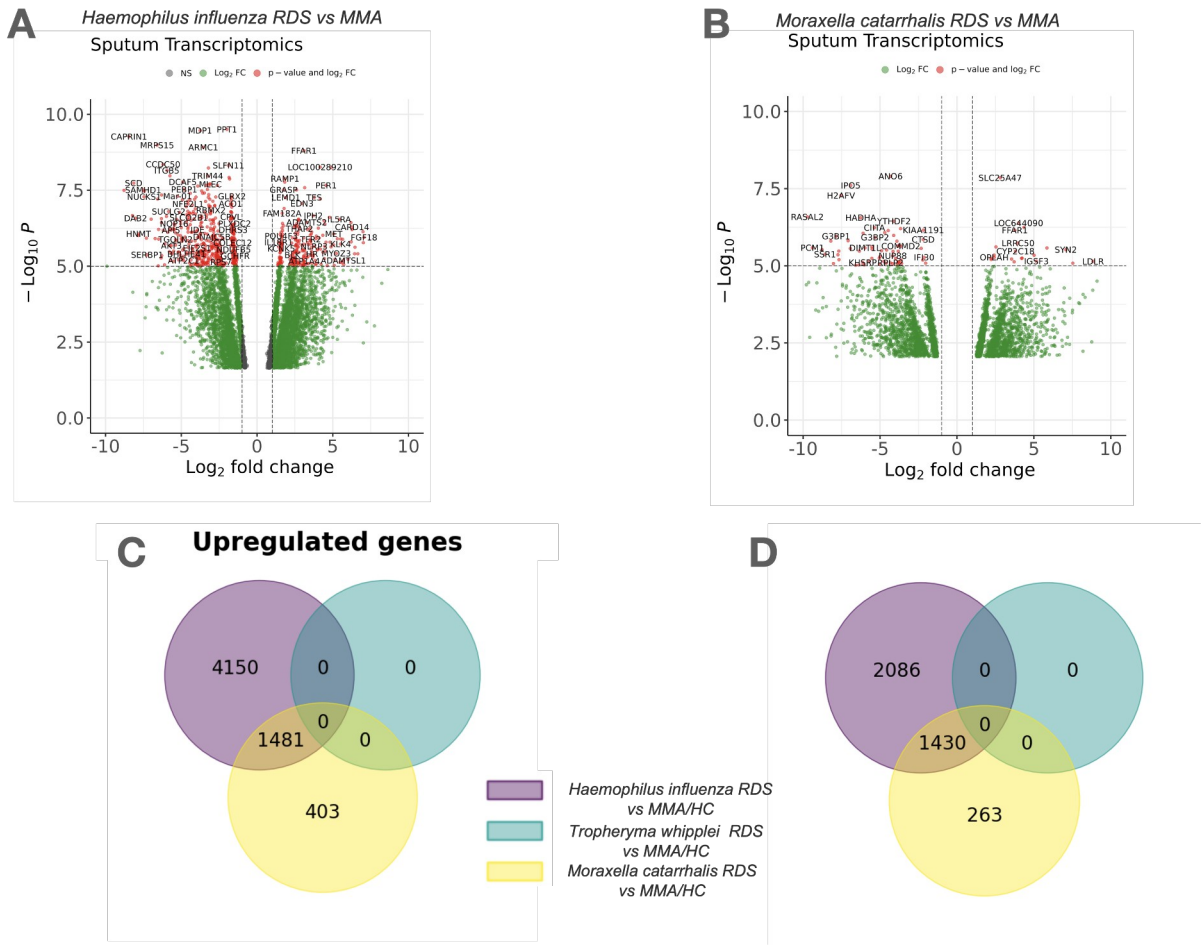
In 51% of samples with low diversity, *Haemophilus influenzae* was not identified. Patients with *Haemophilus influenzae* RDS had the longest duration of disease with evidence of more severe disease with highest number of exacerbations in the previous year and of patients on oral corticosteroid therapy amongst these 3 RDSs (**Table 8**).

They had the second highest sputum neutrophil count after the *Moraxella catarrhalis* RDS ($p<0.05$ compared with non-RDSs).

Relative to the patients with RDSs, sputum transcriptomics identified few DEGs differentially-expressed between severe asthma patients with normal diversity and MMA groups. Patients with *Haemophilus influenzae* RDS have 6,964 more genes upregulated compared to MMA/HC and non-*Haemophilus influenzae* RDSs (

A). Pathway analysis of DEGs showed enrichment of positive regulation of natural killer cell mediated cytotoxicity, positive regulation of type 2a hypersensitivity, antigen processing and toll-like receptor signalling pathway compared to the MMA group (**Figure 34A**).

Figure 33. Transcriptomic differentially-expressed genes between RDS groups and MMA/HC cohort.



Legend: Volcano plot depicting the sputum transcriptomic DEG between (A) Haemophilus influenzae RDS vs MMA/HC, (B) Moraxella catarrhalis RDS vs MMA/HC, (C) Tropheryma whipplei RDS vs MMA/HC. Venn diagram depicts the overlap between up-regulated genes (D) and down-regulated genes (E) from the differentially-expressed genes of the sputum transcriptomics between the Haemophilus influenzae, Moraxella catarrhalis and Tropheryma whipplei RDSs and MMA/HC. MMA/HC: mild-moderate asthma and healthy controls.

5.3.2.2 *Moraxella catarrhalis RDS*

Moraxella catarrhalis was detectable in 10% (10/97) of severe asthmatic (SAn & SAs/ex) samples with a mean abundance of 6.79% ± 23.3% (interquartile range, IQR) and 40% (4/10) of samples met the loss of diversity criteria of RDS. The mean

abundance of *Moraxella catarrhalis* within RDS samples was 24.44 ± 41.35. Patients with *Moraxella catarrhalis*

RDS had the shortest duration of disease but had the highest sputum neutrophils (**Table 8**)**.** Sputum eosinophils were low at 1.4% (p=0.001 compared with non-RDSs comparator) and FeNO was also reduced (p<0.05). In addition, these subjects had less exacerbations per year (p<0.05 compared with non-RDSs). Patients with *Moraxella catarrhalis* RDS had 2,820 genes upregulated compared to MMA *(*

B &

D). Gene ontology analysis of DEGs showed positive regulation of diacylglycerol kinase activity and granulocyte colony-stimulating factor and macrophage colony stimulating factor production and negative regulation of LPS-mediated signalling pathway and IL-13 production compared to MMA/HC (**Figure 34B**)

Table 8. Characteristics of severe asthma with Haemophilus influenzae (Hi), Moraxella catarrhalis (Mc) & Tropheryma whipplei (Tw) relative dominant species (RDS).

	Hi RDS	*Mc* RDS	*Tw* RDS	non-RDSs	MMA/HC	P Value	#P value	##P Value
Subjects, n	15	4	5	46	48	NA	NA	NA
Age (years)	51.0 [45.5,56.5]	57.5 [48.2,62.0]	62.0 [46.0,71.0]	55.0 [45.2,62.8]	42.0 [27.8,50.2]	<0.001	NS	NS
Females, n (%)	9 (60.0)	3 (75.0)	3 (60.0)	29 (63.0)	19 (39.6)	NS	NS	NS
BMI	26.4 (2.6)	27.7 (4.5)	29.9 (5.0)	28.8 (6.3)	25.6 (4.1)	<0.05	NS	NS
Duration years,	38.2 (14.0)	18.0 (21.5)	25.0 (13.0)	22.4 (16.1)	23.5 (16.6)	<0.05	<0.05	<0.05
Current Smoking, n (%)	2 (13.3)	0 (0)	0 (0)	6 (13.0)	0 (0)	<0.005	NS	NS
Oral corticosteroid daily, n (%)	7 (46.7)	1 (25.0)	1 (20.0)	17 (37.0)	0 (0)	<0.001	NS	Ns
ACQ-5 score,	2.2 (1.3)	2.9 (1.6)	2.8 (1.1)	2.3 (1.2)	0.9 (0.8)	<0.001	NS	NS
Exacerbations, per year,	2.9 (2.5)	1.0 (0.8)	1.2 (1.3)	2.4 (2.1)	0.3 (0.8)	<0.001	NS	NS

Nasal polyposis, n (%)	4 (26.7)	2 (50.0)	2 (40.0)	15 (32.6)	6 (12.5)	NS	NS	NS
Eczema, n (%)	6 (40.0)	1 (25.0)	0 (0)	11 (23.9)	9 (18.8)	NS	NS	NS
Allergic rhinitis, n (%)	8 (53.3)	3 (75.0)	1 (20.0)	11 (23.9)	15 (31.2)	NS	NS	NS
GERD, n (%)	6 (40.0)	1 (25.0)	2 (40.0)	18 (39.1)	4 (8.3)	<0.005	NS	NS
Blood Eosinophils μL^{-1}	0.3 (0.2)	0.5 (0.6)	0.4 (0.2)	0.3 (0.3)	0.2 (0.1)	0.005	NS	NS
Blood Neutrophils μL^{-1}	6.5 (2.5)	6.8 (4.4)	4.8 (1.5)	5.1 (2.6)	3.6 (1.3)	<0.001	NS	NS
Sputum Neutrophils (%),	68.0 (27.7)	82.1 (25.4)	56.2 (11.0)	49.5 (24.0)	45.5 (26.1)	<0.005	<0.05	NS
Sputum Eosinophils (%)	19.4 (25.2)	1.4 (1.8)	19.7 (13.2)	13.1 (20.5)	2.1 (8.1)	<0.001	NS	NS
Sputum Lymphocytes (%)	0.8 (0.6)	0.7 (0.5)	1.2 (1.3)	1.6 (1.4)	1.9 (2.0)	NS	NS	NS
Sputum Macrophage (%)	11.8 (8.3)	15.8 (23.2)	22.8 (18.8)	35.7 (22.5)	50.9 (26.0)	<0.001	0.001	NS
FEV1 (% predicted)	65.1 (26.9)	62.2 (21.1)	40.3 (10.2)	67.4 (18.7)	97.0 (17.2)	<0.001	NS	NS
FeNO (ppb)	30.8 (28.7)	22.2 (10.9)	28.2 (20.1)	40.1 (38.6)	31.0 (26.7)	NS	NS	NS

Smoking pack years	16.8 (11.4)	0 (0)	5.7 (3.3)	18.3 (18.4)	3.2 (1.7)	NS	NS	NS
Antibiotic current, n (%)	2 (13.3)	0 (0)	1 (20.0)	10 (21.7)	0 (0)	<0.05	NS	NS
CRP (mg/L)	10.0 (16.6)	3.6 (1.5)	2.7 (1.8)	6.6 (15.3)	1.3 (1.8)	NS	NS	NS
IL-1α (pg/ml)	34.5 (10.7)	34.8 (6.2)	36.7 (8.3)	35.1 (6.4)	32.3 (5.3)	NS	NS	NS
IL-6 (pg/ml)	1.3 (1.3)	1.6 (0.6)	1.1 (0.4)	1.1 (0.8)	0.6 (0.5)	<0.005	NS	NS
Serum IL-8 (pg/ml)	7.6 (12.7)	3.0 (1.4)	3.1 (0.6)	4.0 (3.0)	3.2 (1.6)	NS	NS	NS
Serum C5a (pg/ml)	60.9 (27.3)	36.7 (9.9)	51.5 (76.0)	39.8 (19.6)	33.8 (19.7)	<0.01	NS	NS

Only samples with no co-occurring RDSs were selected.

Abbreviations: ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; CRP: Serum C-Reactive Protein; FeNO: Fractional exhaled nitric oxide; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; GERD: Gastroesophageal reflux disease; MMA/HC: Mild-moderate asthma/Healthy controls; N/A: Not applicable; NS: Not significant; ppb: parts per billion. Data shown as mean (Standard Deviation) unless variable is categorical where n (%). Age shown as median [IQR].

P value: comparing RDSs and non-RDSs.

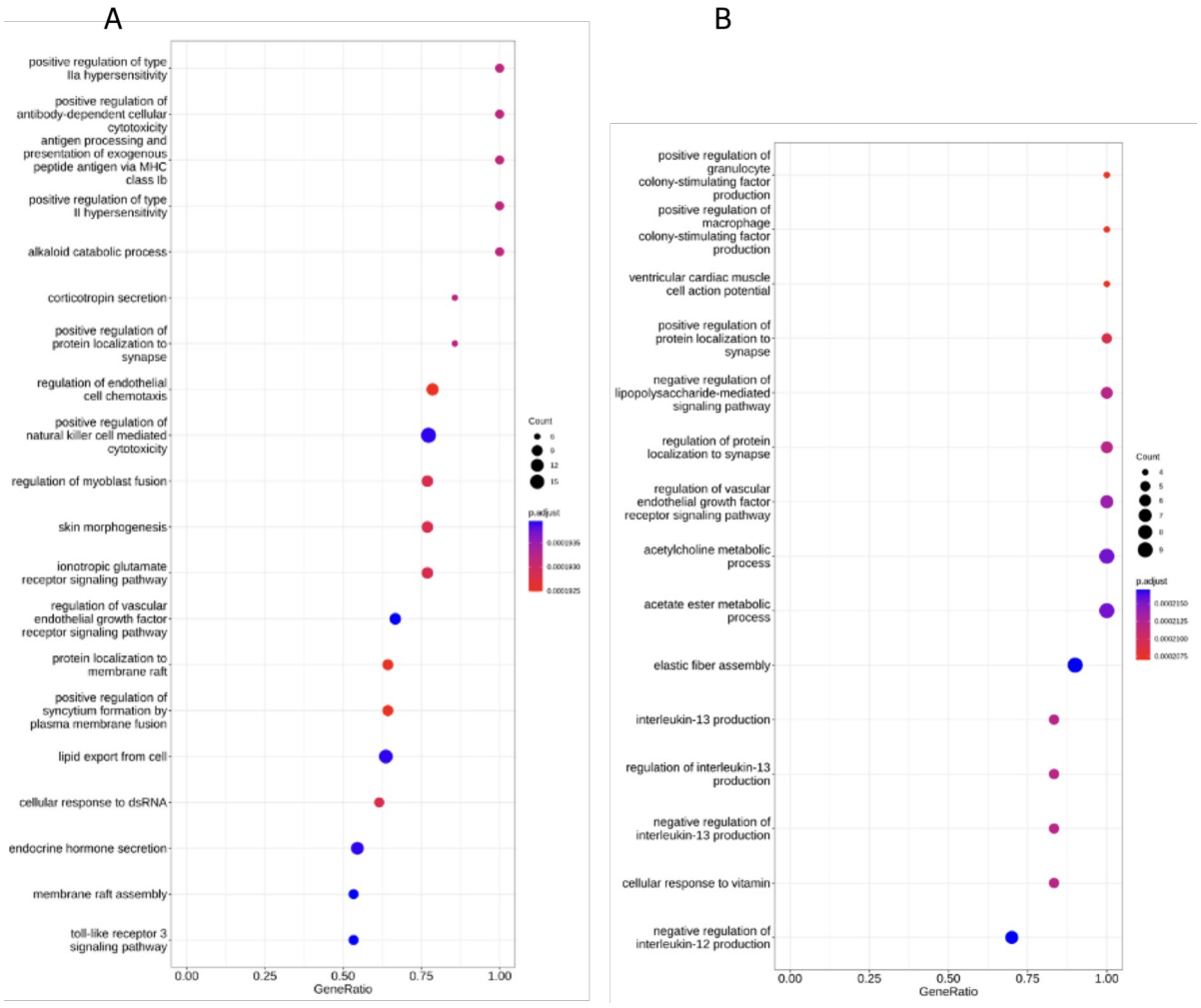
Hi – Haemophilus influenzae, Mc- *Moraxella catarrhalis* and Tw - *Tropheryma whipplei* RDS group

P value: comparing the five groups: Hi RDS, Mc RDS, Tw RDS, non-RDSs, MMA/HC. [#]P value: comparing the four groups: Hi RDS, Mc RDS, Tw RDS, non-RDSs. [##]P value: comparing the three cluster groups: Hi RDS, Mc RDS, Tw RDS.

5.3.2.3 Tropheryma whipplei RDS

*Tropheryma whipplei*_was detected in 14% (14/97) of severe asthmatic (SAn & SAs/ex) samples with abundance of 9.50% ± 48.6%. 36% (5/14) of samples with *Tropheryma whipplei* were RDSs. The mean abundance of *Tropheryma whipplei* within RDS samples was 58.17 ± 12.31. Sputum neutrophils and eosinophils were not different in *Tropheryma whipplei* RDS patients compared to non-RDS subjects (**Table 8**). FEV1 was reduced in *Tropheryma whipplei* RDS (p=0.001 compared with non-RDSs). Patients with *Tropheryma whipplei* RDS had only 4 genes upregulated and none down-regulated compared with MMA.

Figure 34. GO enrichment dot-plots between the Haemophilus influenzae (panel A) and Moraxella catarrhalis relative dominant species (**panel B) compared to MMA group.**

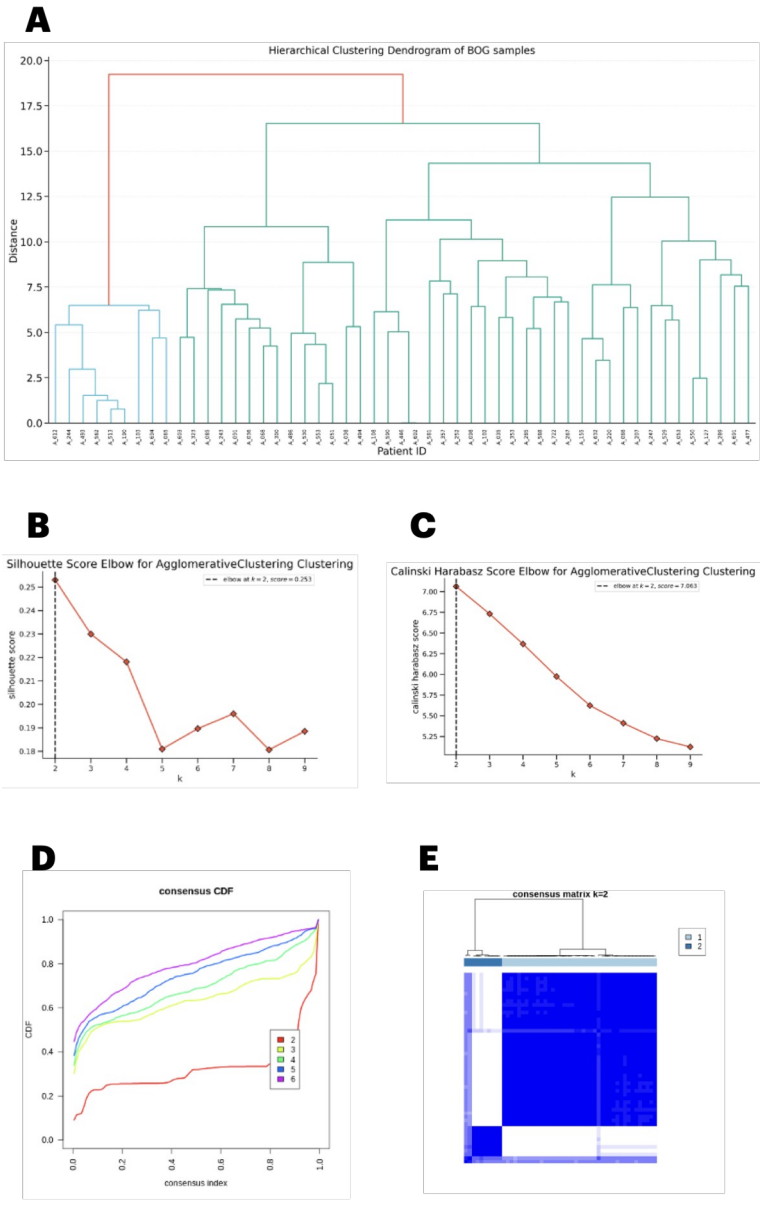


Legend: Dot plots of the top GO enrichment terms (A) Go enrichment dotplot from the DEG of *Haemophilus influenzae* RDS vs MMA (B) Go enrichment dotplot from the DEG of *Moraxella catarrhalis* RDS vs MMA.

5.3.3 Cluster analysis of low diversity samples

Cluster analysis of samples with low diversity identified two clusters as demonstrated on silhouette width (C1 and C2) (**Figure 35B**) and by Calinski-Harabasz score (**Figure 35C**).

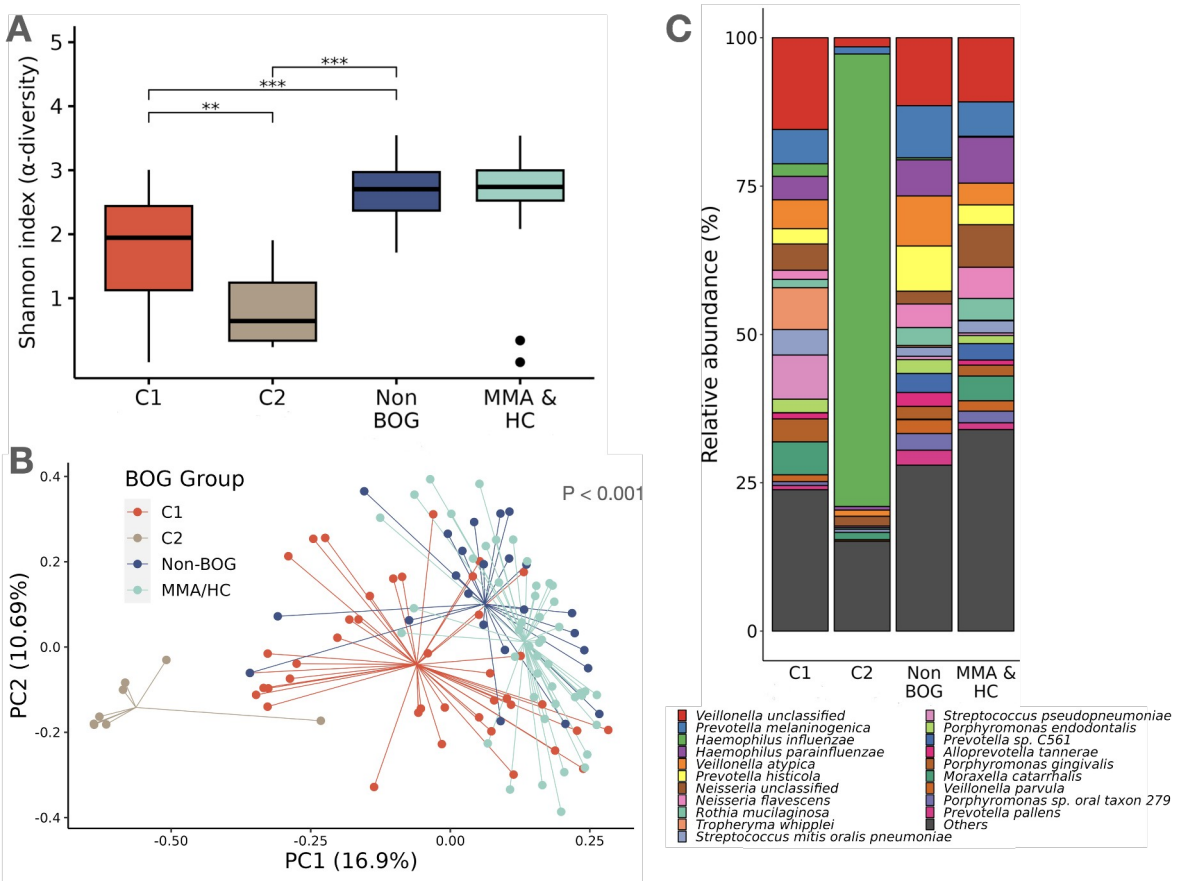
Figure 35. Metagenomic hierarchical clustering of sputum metagenomic species-level abundance of relative dominant specie samples (k=2).



Legend: (A) Dendrogram of the hierarchical clustering. (B) Silhouette score. (C) Calinski-Harabasz score (D) CDF plot displaying consensus distributions for each number of clusters (k). (E) Consensus matrix heat map depicting consensus values on a white to blue colour scale of cluster for κ=2.

Consensus clustering also identified stability with 2 clusters (**Figure 35D and E**) with C2 having the lowest Shannon a-diversity (**Figure 36A**), with the greatest degree of separation shown on the principal component analysis plot (**Figure 36B**).

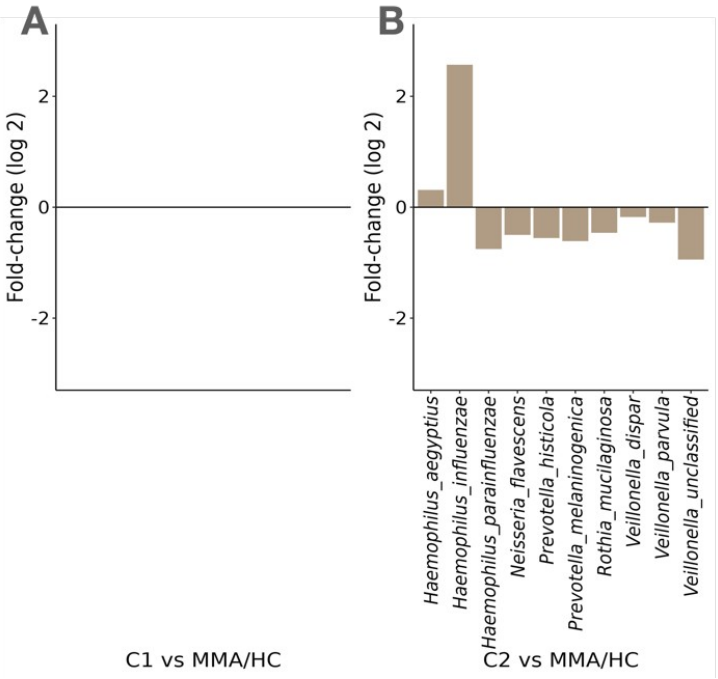
Figure 36. Diversity of Aitcheson distance clusters of RDS samples and Bacterial Species composition.



Legend: α-diversity using Shannon's index of each of the clusters C1 and C2 with Non-RDS and MMA/HC groups (A). β-diversity of the clusters with Non-RDS and MMA/HC groups (B). Species-level relative abundance of each of the 2 clusters compared to that of Non-RDS and MMA/HC groups (C).

C1 had the highest a-diversity with greatest similarity to MMA/HC patients, corroborated by the differential abundance analysis with minimal changes in species abundance (**Figure 37A & Figure 37B**).

Figure 37. Composition log-fold change between each of the C-Hi (A) and C2-Hi (B) clusters and the MMA/HC group using ANCOM-BC



Legend: Log2-fold change of the composition using ANCOM-BC between C1-Hi and MMA/HC and C2-Hi and MMA/HC.

C2 was characterised by a very large increased abundance of *Haemophilus influenzae* with an abundance of 76.2±19.9% compared to 2.1 ± 5.1% for C1 (**Table 9** and **Figure 36C**). By contrast, *Tropheryma whipplei* and *Moraxella catarrhalis* were more abundant in C1 than in C2. While C1 had no differentially-abundant bacterial species difference with MMA/HC (**Figure 37A** and **Table 9**), C2 had more than 2-fold higher abundance of *Haemophilus influenzae* compared to MMA/HC, with reduced other bacterial species including *Haemophilus parainfluenzae*.

Table 9. Abundance of RDS species present in C1 and C2 clusters and in MMA/HC

Bacterial species BOG	C1 (n=42)	C2 (n=9)	Non-RDS (n=46)
Tropheryma whipplei	7.06 ± 17.92 (R=5)	0.00 ± 0.00 (R=0)	0.29 ± 1.88
Corynebacterium pseudodiphtheriticum	2.38 ± 15.43 (R=1)	0.00 ± 0.00 (R=0)	0.00 ± 0.00
Rothia dentocariosa	1.70 ± 6.26 (R=2)	0.09 ± 0.22 (R=0)	0.81 ± 1.50
Propionibacterium acnes	2.39 ± 9.77 (R=5)	0.07 ± 0.21 (R=0)	0.01 ± 0.06
Porphyromonas endodontal	2.27 ± 10.33 (R=1)	0.00 ± 0.00 (R=0)	2.34 ± 6.09
Porphyromonas gingival	3.86 ± 13.79 (R=1)	0.00 ± 0.00 (R=0)	2.15 ± 5.79
Alloprevotella unclassified	1.01 ± 5.00 (R=1)	0.00 ± 0.00 (R=0)	0.59 ± 1.63
Prevotella baroniae	0.32 ± 1.96 (R=1)	0.00 ± 0.00 (R=0)	0.20 ± 0.97
Prevotella dentalis	0.09 ± 0.48 (R=1)	0.00 ± 0.00 (R=0)	0.07 ± 0.20
Deinococcus unclassified	0.42 ± 2.71 (R=1)	0.00 ± 0.00 (R=0)	0.03 ± 0.09
Streptococcus pneumoniae	4.29 ± 8.47 (R=4)	0.51 ± 1.08 (R=0)	1.52 ± 2.16
Streptococcus pseudopneumoniae	7.46 ± 16.97 (R=7)	0.00 ± 0.00 (R=0)	0.57 ± 1.03
Veillonella unclassified	15.47 ± 23.76 (R=9)	1.53 ± 3.22	11.46 ± 6.50

		(R=0)	
Fusobacterium nucleatum	0.41 ± 1.90 (R=1)	0.00 ± 0.00 (R=0)	0.39 ± 0.79
Comamonas unclassified	0.04 ± 0.24 (R=1)	0.00 ± 0.00 (R=0)	0.00 ± 0.00
Kingella unclassified	0.80 ± 4.75 (R=1)	0.30 ± 0.80 (R=0)	0.37 ± 0.75
Neisseria unclassified	4.44 ± 9.50 (R=2)	1.66 ± 2.88 (R=0)	2.15 ± 3.19
Actinobacillus unclassified	0.81 ± 2.75 (R=4)	3.68 ± 1.79 (R=6)	0.54 ± 0.69
Haemophilus aegyptius	0.17 ± 0.79 (R=2)	2.62 ± 2.42 (R=7)	0.07 ± 0.21
Haemophilus influenzae	2.12 ± 5.13 (R=7)	76.28 ± 19.92 (R=9)	0.34 ± 0.73
Haemophilus parainfluenzae	3.94 ± 8.08 (R=1)	0.59 ± 1.10 (R=0)	6.09 ± 4.89
Moraxella catarrhalis	5.58 ± 20.63 (R=4)	1.24 ± 3.71 (R=1)	0.09 ± 0.39
Pseudomonas aeruginosa	0.00 ± 0.00 (R=0)	2.65 ± 7.94 (R=1)	0.00 ± 0.00
Pseudomonas unclassified	0.00 ± 0.00 (R=0)	0.47 ± 1.41 (R=1)	0.00 ± 0.00
Eimeria tenella	0.03 ± 0.18 (R=1)	0.00 ± 0.00 (R=0)	0.00 ± 0.00

Abbreviations:

RDS: Relative dominant specie; MMA/HC: Mild-moderate asthma/Healthy controls; Data shown as mean ± Standard Deviation with number of RDS samples for that species in parenthesis. C1 – Cluster 1 of RDS samples; C2 – Cluster 2 of RDS samples. B denotes the number of samples which contain RDSs for that species.

Haemophilus influenzae-dominant C2 cluster had a longer disease duration compared to C1 (41.6 ± 12.4 vs 24.9 ± 15.3; p ≤ 0.005) (**Table 10**). Sputum neutrophil counts (%) were also higher in C2 (85.5 ± 13.5% vs 58.9 ± 24.2%; p< 0.001). Sputum eosinophil counts (%) on the other hand were higher in the C1 (16.4 ± 21.4 vs 5.6 ± 10.4; P ≤ 0.05) and sputum macrophages (%) were lower in C2 (p< 0.001). Serum IL8 was higher in C2 compared to C1 and non-RDS groups (p<0.05). Gene ontology pathway analysis of DEGs of sputum between C1 and C2 clusters showed enrichment of pathways in C2 over C1 related to MAPK, NF-kB, TNF, mTOR and necroptosis (

).

Table 10. Characteristics of clusters C1 and C2, and non-RDSs

	C1	C2	non-RDSs	#P value*	##P value
Subjects, n	42	9	46	ND	ND
Age (years)	54.5 [47.0,59.8]	53.0 [46.0,55.0]	55.0 [45.2,62.8]	NS	NS
Females, n (%)	21 (50.0)	6 (66.7)	29 (63.0)	NS	NS
BMI	28.7 (5.2)	28.0 (4.5)	28.8 (6.3)	NS	NS
Duration (years)	24.9 (15.3)	41.6 (12.4)	22.4 (16.1)	<0.005	<0.005
Current Smoking, n (%)	4 (9.5)	1 (11.1)	6 (13.0)	NS	NS
Oral corticosteroid daily, n (%)	18 (42.9)	4 (44.4)	17 (37.0)	NS	NS

ACQ-5 score	2.3 (1.3)	2.3 (1.1)	2.3 (1.2)	NS	NS
Exacerbations, per year	1.8 (2.1)	3.4 (2.2)	2.4 (2.1)	<0.05	NS
Nasal polyposis, n (%)	16 (38.1)	2 (22.2)	15 (32.6)	NS	NS
Eczema, n (%)	9 (21.4)	5 (55.6)	11 (23.9)	NS	NS
Allergic rhinitis, n (%)	20 (47.6)	5 (55.6)	11 (23.9)	<0.05	NS
GERD, n (%)	19 (45.2)	5 (55.6)	18 (39.1)	NS	NS
Blood Eosinophils μL^{-1}	0.3 (0.3)	0.3 (0.2)	0.3 (0.3)	NS	NS
Blood Neutrophils μL^{-1}	5.5 (2.4)	6.8 (2.7)	5.1 (2.6)	NS	NS
Sputum Neutrophils (%)	58.9 (24.2)	85.5 (13.5)	49.5 (24.0)	<0.001	<0.001
Sputum Eosinophils (%)	16.4 (21.4)	5.6 (10.4)	13.1 (20.5)	NS	<0.05
Sputum Lymphocytes (%)	1.1 (1.1)	0.6 (0.4)	1.6 (1.4)	<0.05	<0.05
Sputum Macrophage (%)	23.6 (19.1)	8.2 (7.9)	35.7 (22.5)	<0.001	0.001
FEV1 (% predicted)	63.9 (20.1)	55.6 (26.5)	67.4 (18.7)	NS	NS
FeNO (ppb)	38.3 (37.6)	19.8 (8.7)	40.1 (38.6)	NS	<0.01
Smoking pack years	15.2 (16.7)	15.9 (12.9)	18.3 (18.4)	NS	NS
Antibiotic current, n (%)	5 (11.9)	2 (22.2)	10 (21.7)	NS	NS
Serum CRP (mg/L)	6.1 (12.5)	9.1 (10.5)	6.6 (15.3)	NS	NS
Serum IL-1α (pg/ml)	36.7 (6.3)	31.3 (12.6)	35.1 (6.4)	NS	NS
Serum IL-6 (pg/ml)	1.4 (1.1)	1.5 (1.5)	1.1 (0.8)	NS	NS
Serum IL-8 (pg/ml)	3.7 (1.4)	9.6 (15.8)	4.0 (3.0)	<0.05	NS

Serum C5a (pg/ml)	49.0 (30.5)	56.7 (18.9)	39.8 (19.6)	NS	NS

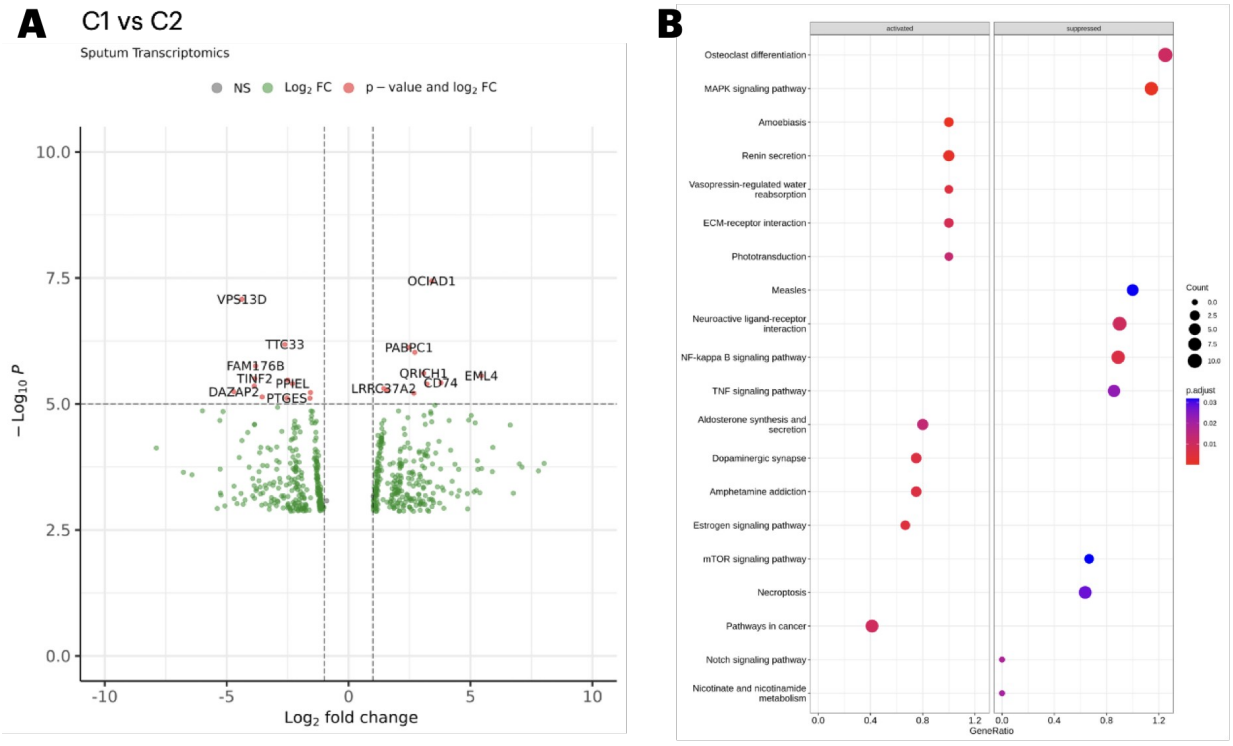
Abbreviations:

ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; CRP: Serum C-Reactive Protein; FeNO: Fractional exhaled nitric oxide; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; GERD: Gastroesophageal reflux disease; N/A: Not applicable; NS: Not significant; ppb: parts per billion. Data shown as mean (Standard Deviation) unless variable is categorical where n (%). Age shown as median [IQR]. P value: comparing RDSs and non-RDSs.

#P value: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the three groups: C1-RDS, C2-RDS, non-RDSs.

##P value: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the two cluster groups: C1-RDS, C2-RDS.

Figure 38. DEG and pathway analysis of clusters C1 vs C2



Legend: Panel A. DEG and GO pathway enrichment between C1 and C2 clusters. Panel B. Volcano plot depicts the sputum transcriptomics DEG between C1 and C2 cluster. Dot-plots depict the GO pathway enrichment for C1 and C2 cluster.

5.3.4 Haemophilus influenzae RDSs of C1 and C2 clusters

Haemophilus influenzae RDS samples were composed of 7/42 in C1 (denoted as C1-Hi) and 9/9 in C2 samples (denoted as C2-Hi), that were differentiated by sputum neutrophilic inflammation in C1 and eosinophilic inflammation in C2 (

and

B).

Table 11. Characteristics of severe patients with the Haemophilus influenzae RDSs samples in C1 and C2 clusters

	C1-*Hi* (n=7)	C2-*Hi* (n=9)	Non-*Hi*-RDS	#P Value	##P Value	###P Value
Subjects, n	7	9	35	NA	NA	NA
Age (years)	47.0 [45.0,52.5]	53.0 [46.0,55.0]	56.0 [50.5,62.0]	NS	NS	NS
Females, n (%)	3 (42.9)	6 (66.7)	18 (51.4)	NS	NS	NS
BMI	26.0 (1.7)	28.0 (4.5)	29.2 (5.5)	NS	<0.01	NS
Duration years,	31.7 (14.9)	41.6 (12.4)	23.6 (15.3)	NS	NS	<0.005
Current Smoking, n (%)	1 (14.3)	1 (11.1)	3 (8.6)	NS	NS	NS
Oral corticosteroid daily, n (%)	4 (57.1)	4 (44.4)	14 (40.0)	NS	NS	NS
ACQ-5 score,	1.9 (1.5)	2.3 (1.1)	2.4 (1.2)	NS	NS	NS
Exacerbations (per year)	2.6 (3.0)	3.4 (2.2)	1.7 (1.8)	NS	NS	<0.05
Nasal polyposis, n	3 (42.9)	2 (22.2)	13 (37.1)	NS	NS	NS

(%)						
Eczema, n (%)	2 (28.6)	5 (55.6)	7 (20.0)	NS	NS	NS
Allergic rhinitis, n (%)	3 (42.9)	5 (55.6)	17 (48.6)	NS	<0.005	NS
GERD, n (%)	6 (40.0)	1 (25.0)	17 (48.6)	NS	NS	NS
Blood Eosinophils μL^{-1}	0.2 (0.2)	0.3 (0.2)	0.4 (0.3)	NS	<0.05	NS
Blood Neutrophils μL^{-1}	6.2 (2.2)	6.8 (2.7)	5.4 (2.4)	NS	NS	NS
Sputum Neutrophils (%),	49.3 (28.2)	85.5 (13.5)	60.9 (23.2)	<0.05	NS	<0.001
Sputum Eosinophils (%)	34.4 (29.3)	5.6 (10.4)	12.8 (17.9)	<0.05	NS	NS
Sputum Lymphocytes (%)	1.1 (0.6)	0.6 (0.4)	1.1 (1.2)	NS	NS	<0.05
Sputum Macrophage (%)	15.3 (7.2)	8.2 (7.9)	25.2 (20.3)	NS	<0.05	<0.001
FEV1 (% predicted)	73.4 (25.6)	55.6 (26.5)	62.0 (18.7)	NS	NS	
FeNO (ppb)	44.6 (37.7)	19.8 (8.7)	37.1 (38.0)	NS	NS	<0.05
Smoking pack years	20.8 (10.5)	15.9 (12.9)	14.4 (17.5)	NS	NS	NS

Antibiotic current, n (%)	0 (0)	2 (22.2)	5 (14.3)	NS	NS	NS
CRP (mg/L)	10.7 (22.2)	9.1 (10.5)	5.2 (9.9)	NS	NS	NS
IL1α (pg/ml)	38.0 (5.2)	31.3 (12.6)	36.4 (6.5)	NS	NS	NS
IL6 (pg/ml)	1.0 (0.9)	1.5 (1.5)	1.4 (1.1)	NS	NS	NS
Serum IL8 (pg/ml)	4.2 (1.4)	9.6 (15.8)	3.6 (1.3)	NS	NS	NS
Serum C5a (pg/ml)	65.8 (34.8)	56.7 (18.9)	45.7 (28.9)	NS	NS	NS
H.influenzae abundance (%)	12.2 (6.1)	76.3 (19.9)	0.1 (0.3)	<0.001	<0.005	<0.001

Abbreviations:

ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; CRP: Serum C-Reactive Protein; FeNO: Fractional exhaled nitric oxide; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; GERD: Gastroesophageal reflux disease; MMA/HC: Mild-moderate asthma/Healthy controls; N/A: Not applicable; NS: Not significant; ppb: parts per billion. Data shown as mean (Standard Deviation) unless variable is categorical where n (%). Age shown as median [95% Confidence intervals].

C1-*Hi* – Cluster 1 of *Haemophilus influenzae* RDS; C2-*Hi* – Cluster 2 of *Haemophilus influenzae* RDS; Non-*Hi*-RDS – Non-*Haemophilus influenzae* RDS cluster.

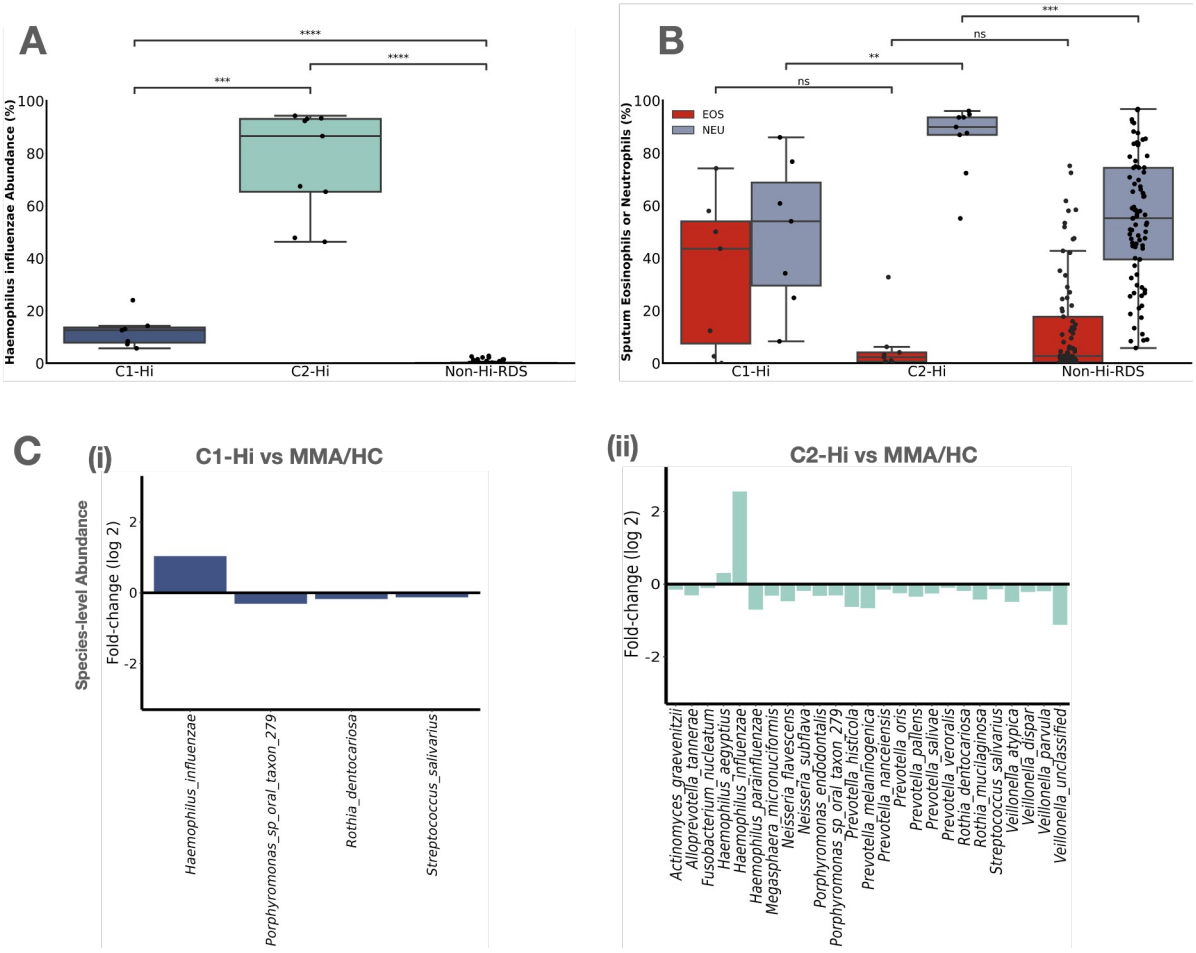
[#]**P Value:** C1-Hi vs C2-Hi; [##]**P Value:** C1-Hi vs Non-Hi-RDS; [###]**P Value:** C2-Hi vs Non-Hi-RDS;

The relative abundance of *Haemophilus influenzae* in C2-*Hi* (76.3 ± 19.9%) was higher than in C1-*Hi* (12.2 ± 6.1%) (

A), and when compared to MMA/HC, C2-*Hi* had more than 4-fold higher abundance of *Haemophilus influenzae* compared to MMA/HC, with reduced other bacterial species including *Haemophilus parainfluenzae* (

C). C1-*Hi* had 2-fold higher abundance than MMA/HC, with no other difference of other bacterial species.

Figure 39. Haemophilus influenzae component of C1 (C1-Hi) and C2 (C2-Hi) clusters of RDS samples.

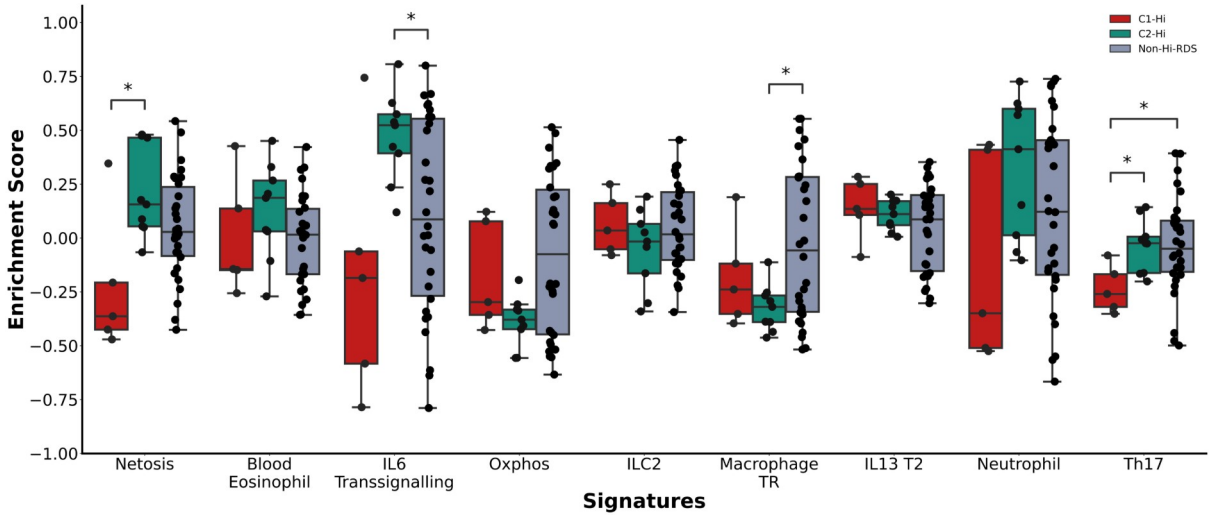


Legend: **A:** Haemophilus influenzae abundance of C1-Hi and C2-Hi with Non-Hi-RDS. **B:** Sputum eosinophils and sputum neutrophils (%) of C1-Hi and C2-Hi with Non-Hi-RDS. **C:** Composition log-fold change between each of the C1-Hi (i) and C2-Hi (ii) clusters and the MMA/HC group. RDS: Relative dominant species of severe asthma groups; MMA/HC: mild-moderate asthma and healthy controls. C1-Hi (Haemophilus influenzae) component of Cluster 1 RDS samples; C2-Hi (Haemophilus influenzae) component of Cluster 2 RDS samples.

Using gene set variation analysis, there was a higher expression score of the gene signature of NETosis (256) and of Th17 activation (257) in the C2-Hi cluster compared to C1-Hi, while for the IL6-transignalling (257), the score was also higher in C2 but not significantly because of a high score in one C1-Hi subject (

). There were no significant differences for the gene signatures representing blood eosinophils, neutrophil activation, oxidative phosphorylation, innate lymphoid cell type-2 (ILC2) and lung-tissue resident macrophages.

Figure 40. Gene set variation analysis (GSVA) enrichment score between the C1 and C2 cluster component of Haemophilus Influenzae RDS (C1-Hi & C2-Hi) and the Non-Hi-RDS.



Legend: Boxplot of the enrichment scores of 9 gene signatures representing NETosis [**SIG6**] (256,258), Blood Eosinophil [**SIG8**] (259), IL6-transignalling [**SIG7**] (260), Oxidative phosphorylation (Oxphos) [**SIG5**] (196), Innate lymphoid cell type 2 (ILC2) [**SIG3**] (193), Lung-Tissue Resident Macrophage (Macrophage TR) [**SIG1**] (189), IL13-T2 [**SIG4]** (192), Neutrophil [**SIG2**] (194) and Th17 [**SIG9**] (257). GSVA was performed in R using the

Bioconductor GSVA package to estimate the variation in gene signature enrichment scores (ES).

Mann–Whitney U test was used for comparison between groups. Significance: *:P≤0.05.

On the other hand, there were no DEGs of sputum transcriptomics between C1-*Hi* and C2-*Hi,* and between C1-*Hi* and Non-*Hi*-RDS groups (

A,

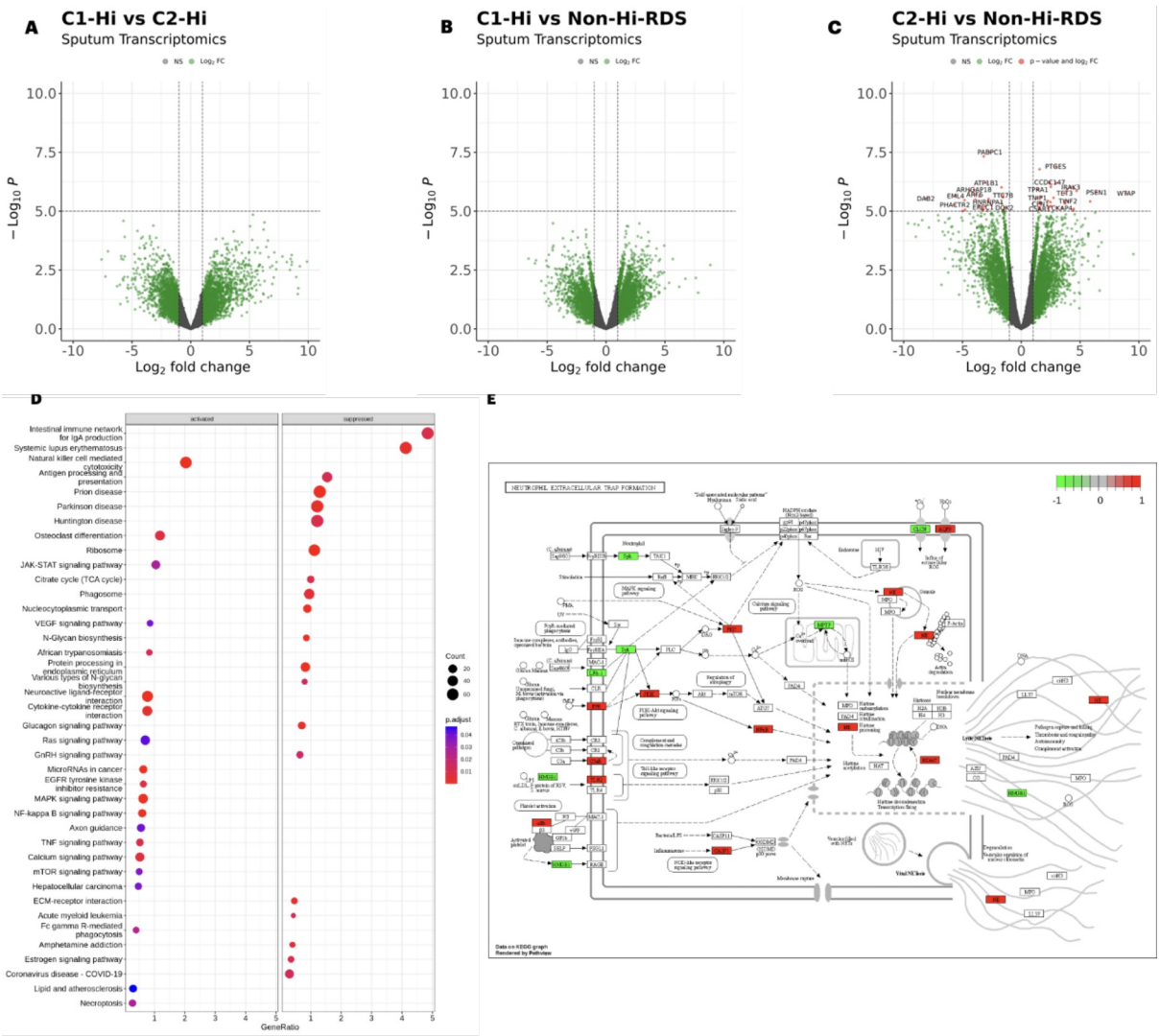
B &

C). However, there were 2,944 DEGs between C2-*Hi* and Non-*Hi*-RDS, particularly with the neutrophil extracellular trap formation being more active in C2-*Hi* than in non-*Hi*-RDS (

D &

E), which is confirmed by the GSVA analysis.

Figure 41. Comparison of C1-Hi, C2-Hi and Non-Hi-RDS groups

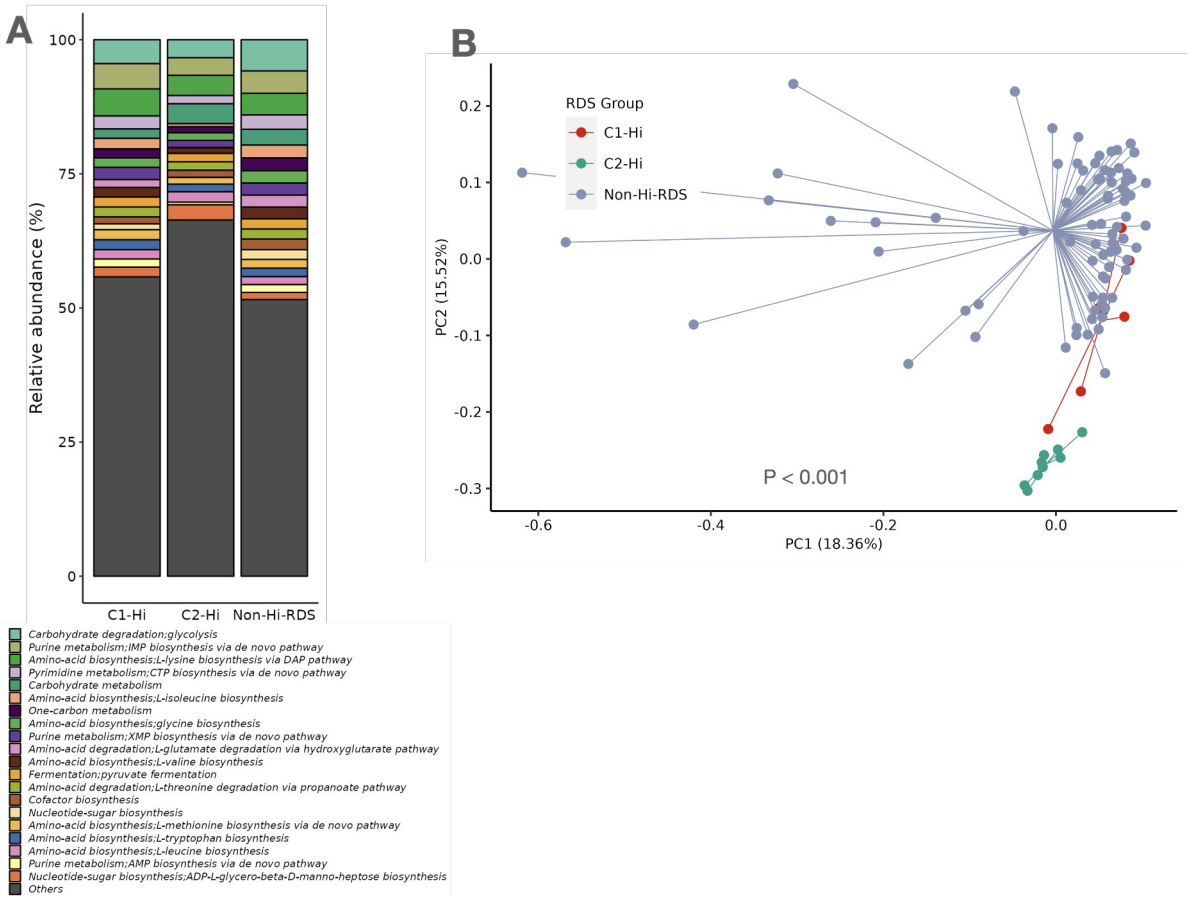


Legend: Comparison of C1-Hi, C2-Hi and Non-Hi-RDS groups: (A) Differentially-expressed genes (DEGs) in sputum between C1-Hi and C2-Hi groups. (B) DEGs between C1-Hi and Non-Hi-RDS groups. (C) DEGs between C2-Hi and Non-Hi-RDS groups. (D) KEGG enrichment pathway dot-plot between C2-Hi and Non-Hi-RDS groups. (E) KEGG pathway for NETosis with C2-Hi and Non-Hi-RDS groups DEG overlayed.

A distinct microbial pathway profile was found between C1-*Hi* and C2-*Hi* by metagenomic functional analysis (**Figure 42A** & **Figure 42B**) with difference in PERMANOVA of the Bray-Curtis dissimilarity matrix ($p < 0.001$). 2 significantly-different UniProt sub-pathways were found in C2-Hi compared to C1-*Hi* when applying differentially-abundant pathway analysis using ANCOM-BC (

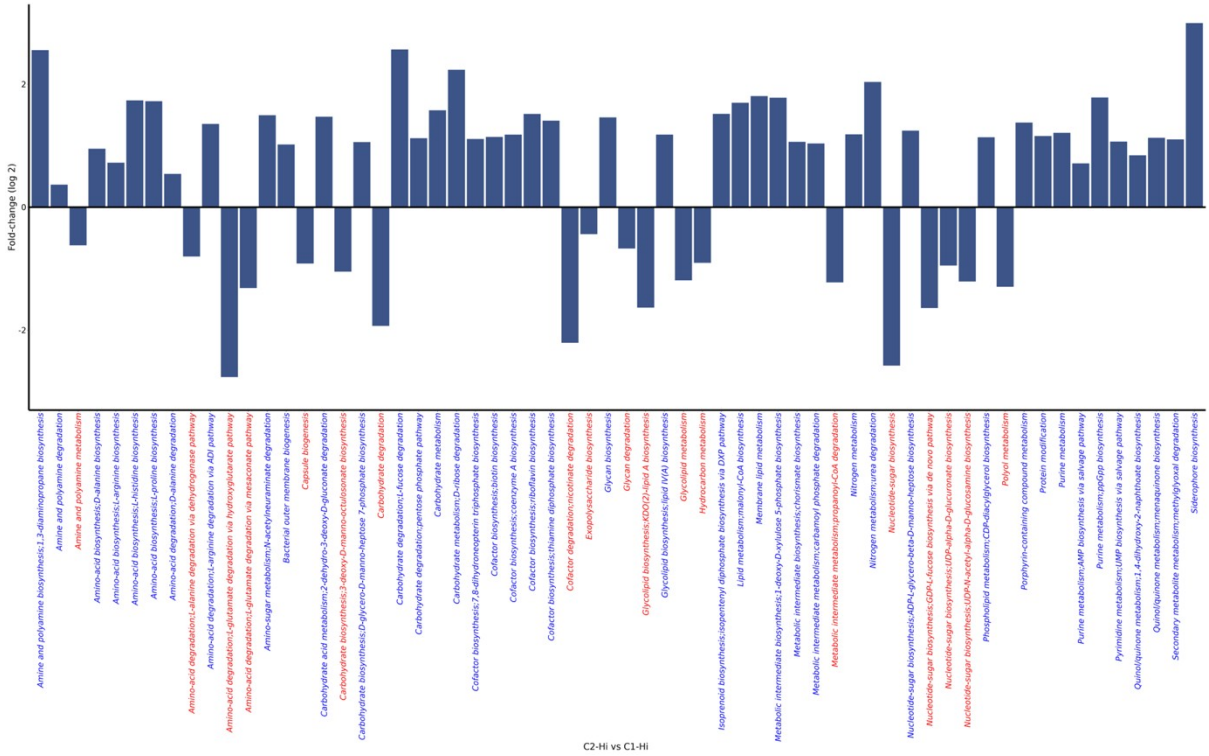
). Further analysis showed that pathways of carbohydrate metabolism and biosynthesis associated with glycolysis are more abundance in C2-*Hi* compared to C1-*Hi* of.

Figure 42. Microbial pathway prevalence in C1-Hi and C2-Hi cluster samples



Legend: (A) Microbial pathway relative abundance of each of the 2 Haemophilus influenzae clusters (C1-Hi and C2-Hi) compared to that of Non-Hi-RDS. (B) β-diversity of the relative abundant microbial pathways of the Haemophilus influenzae clusters (C1-Hi and C2-Hi) with Non-Hi-RDS.

Figure 43. ANCOM-BC plot depicting differentially prevalent microbial pathways between C2-Hi and C1-Hi.

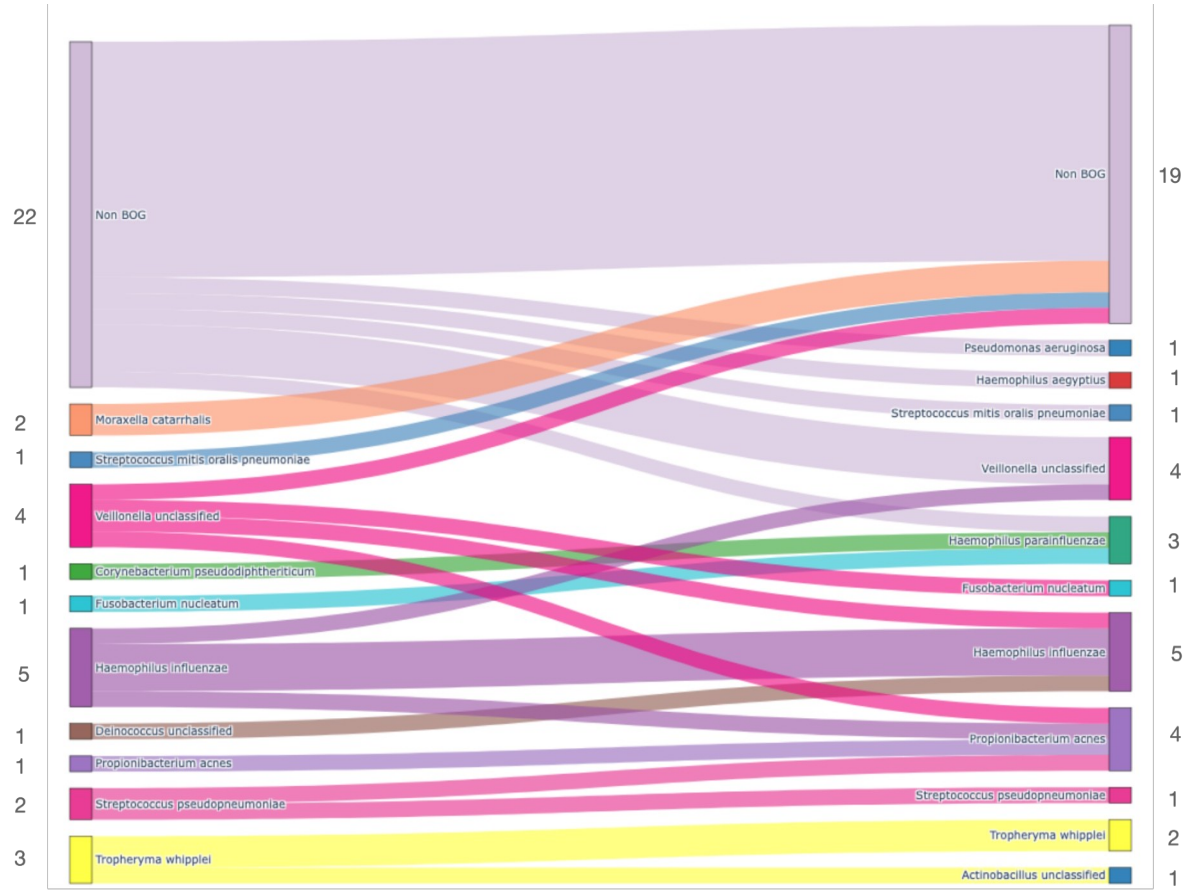


Legend: The blue labels indicate a positive log-fold change and the red labels indicate a negative log-fold change.

5.3.5 *Longitudinal Analysis of severe asthma samples*

At one year follow-up, 7 of the 22 non-RDS samples became RDS with dominance of *Pseudomonas aeruginosa*, *Haemophilus aegyptius*, *Haemophilus parainfluenzae*, *Streptococcus mitis oralis pneumoniae* and *Veillonella unclassified*. 13 of 21 RDSs samples changed RDS status in terms of bacterial species with a shift for *Moraxella catarrhalis* (2 out of 2), *Haemophilus influenzae* (2 out of 5) and for *Tropheryma whipplei* (1 out of 3).

Figure 44. Sankey plot of the stability of dominant RDS over time.



Legend: Sankey plot depicting the stability of the dominant RDS within each sample at baseline and follow-up. The number of samples for the dominant RDSs and non-RDS groups (n=43 in total) at baseline and at one year follow-up is shown alongside each of the 2 time-points with regards to the bacterial species

5.4 Discussion

The present study employed a new approach (RDS) to establish which bacterial species were prone to major shifts in abundance in severe asthmatics. The approach consisted of the quantification of species abundance in induced sputum samples from severe asthma patients using the state of significantly-low diversity as a comparator. The latter was defined according to the diversity variance obtained from the control groups of MMA and HC. The RDS approach was deemed appropriate considering the high susceptibility of low-abundance taxa to exhibit major shifts in their growth and abundance (255,256). Moreover, lowly-abundant or 'rare' taxa play a major role in supporting and maintaining ecosystem/community structure and function that is not commensurate with their abundance (257).

Low α-diversity was accompanied by outgrowths of 24 bacterial species, many of which are of unknown pathogenic significance. Of the known pathogenic species with implications in airway diseases such as asthma and COPD, H*aemophilus influenzae* most frequently constituted a RDS (44%) followed by *Moraxella catarrhalis* and *Tropheryma whipplei*, which each contributed to 14.4% of detected RDSs. Upon evaluation, species of RDS had variable associations with some clinical and inflammatory factors. In fact, the highest number of exacerbations, highest daily corticosteroid use, and longest duration of asthma were associated with *Haemophilus influenzae* RDS. The latter was also associated with elevated neutrophils but decreased macrophages in the sputum, with some subjects exhibiting high sputum eosinophilia. As for the *Moraxella catarrhalis* RDS, it was associated with the worst asthma control (highest asthma control questionnaire (ACQ-5) score), as well as sputum neutrophilia, low sputum eosinophils and FeNO. *Tropheryma whipplei* RDS was associated with both increased levels of eosinophils in the sputum and FeNO, which are suggestive of type-2 inflammation and severe airflow obstruction.

The data also supports that the neutrophilic T2-low asthma phenotype is characterized by an increased abundance of *Haemophilus influenzae*. Cluster analysis of bacterial BOGs led to the identification of a small but select cluster (C2)

with the highest abundance of *Haemophilus influenzae* that had a distinct association with longer duration of disease, increased neutrophil levels in the sputum and higher serum IL-8 levels, but decreased counts of sputum macrophages and eosinophils, and low FeNO levels. These findings further supported the association between increased abundance of *Haemophilus influenzae* and both neutrophilic and a low T2 inflammation. Therefore, the C2 cluster membership composed exclusively of *Haemophilus influenzae* RDS and no other RDS represented a neutrophilic phenotype without eosinophilic inflammation. We were thus able to establish, for the first time, an association between the *Haemophilus influenzae* cluster with enrichment of pathways associated with MAPK, NF-κB, TNF, mTOR and necroptosis. This is consistent with previous reports of the cross-talk between *Haemophilus influenzae* and epithelial cells (258,259).

Our results also reveal that the high abundance of *Haemophilus influenzae* likely elicits a host immune response as evidenced by the high expression of the gene signature for neutrophil extracellular traps (NETs) in the *Haemophilus influenzae*-dominant C2-Hi cluster. In response to pro-inflammatory stimuli (e.g. IL-8 and, TNF), and downstream kinases (e.g. MAPK and mTOR) neutrophils are activated leading to the release of NETs, which are web-like scaffolds of DNA in complex with histones and neutrophil granular proteins. While secreted as part of the host immune response to pathogens, excessive NETs in severe asthma may lead to damage of the airway epithelium, which in turn will perpetuate NETopathic inflammation, and ultimately, cause lung tissue injury (260). It is therefore entirely possible that this innate mechanism, namely the release of NET for pathogen entrapment and neutralization, occurs in response to the *Haemophilus influenzae* dominating the microbiome of C2-Hi cluster subjects found in our study (261).

However, this finding gives rise to the ‘NET paradox’; why are respiratory pathogens commonly found to colonize airways of patients with lung diseases when neutrophils are dedicated to their eradication through the release of NETs. The immune evasion mechanisms (generation of redox active enzymes) developed by *Haemophilus influenzae* to bypass the NET neutralization (242), which itself is triggered directly by this pathogen, gives rise to the ‘NET paradox’ which surrounds NETopathic

inflammation and bacterial colonisation of the airways in certain severe asthma phenotypes. Extant evidence from *in vitro* studies suggests that the contribution of *Haemophilus influenzae* to neutrophilic inflammation (262) could be the result of the invasion and persistence of this pathogen within macrophages (263,264). Moreover, a significantly low percentage of sputum macrophages was linked to the *Haemophilus influenzae* RDS in our study. This suggests that the uncontrolled proliferation and outgrowth of the pathogen could be due to the combined effect of the lower number of airway macrophages and the reduced phagocytic capacity of these cells in severe asthma (228). The resulting impairment of effective macrophage-based immune surveillance likely hinders the elimination of *Haemophilus influenzae* thus contributing to its outgrowth, and by extension, airway dysbiosis. Although fewer in number than the *Haemophilus influenzae* RDS, the *Moraxella catarrhalis* RDS was indicative of another neutrophilic phenotype than that described with *Haemophilus influenzae*. In a confirmation of previous studies on the interactions between *Moraxella catarrhalis* and the airway epithelium (265,266) the *Moraxella catarrhalis* RDS was associated with enrichment of immune pathways implicated in the innate host response against pathogens in addition the activation of granulocyte and macrophage colony stimulating factors.

It is also important to remember that other, less abundant species may also contribute to airway inflammation. However, these species are disregarded when characterizing clusters by abundant species, with one such example being *Rothia mucilaginosa* which was lost in the *Haemophilus*-dominant C2 cluster. Studies have reported that *Rothia mucilaginosa* can inhibit the NF-κB pathway, potentially lending it anti-inflammatory effects (267). Another study found significantly low levels of *Rothia mucilaginosa* in children who are at high risk of asthma, but observed an improvement in airway inflammation in mice who were inoculated with this taxa, suggesting the possible bioprotective role of this bacteria (268). It is therefore plausible that inhibition of other protective or anti-inflammatory species could be one of the mechanisms employed by *Haemophilus influenzae* in order to induce and maintain airway inflammation (269). That being said, very little is known about the bacterial species that had low abundance in the C2 cluster (such as *Prevotella* and *Veillonella)* and the clinical relevance of bacterial species RDS remains ambiguous

apart from *Haemophilus influenzae, Moraxella catarrhalis* and *Tropheryma whipplei.* Therefore, elucidating the properties of these bacterial species and delineating their interactions with the airway epithelial cells and host immune response would be of interest.

This study is not without its limitations. We recognize that the number of samples that constituted a RDS was relatively low, partly due to the low number of sputum samples from severe asthma patients. This limitation could also be attributed to the stringency of the RDS definition, although this approach allowed the strict definition of the *Haemophilus influenzae* RDS and the consequent characterization of its 2 subgroups; the first presenting with neutrophilic inflammation with evidence of NETs and the second presenting with eosinophilic inflammation. Associations could also be drawn with other RDSs albeit with an even lower number of samples, such as the *Moraxella catarrhalis* RDSs which was also associated with diminished sputum macrophages but with purely neutrophilic inflammation. Confirmation of these RDSs following an improved definition of the neutrophilic-associated *Haemophilus influenzae* RDS should be undertaken in a greater number of severe asthmatics. That being said, there has been no similar metagenomic analysis that has been performed in severe asthma.

In summary, we have characterised a distinct phenotype of severe asthma with an increased abundance of *Haemophilus influenzae*. These patients have sputum neutrophilia, associated with a host sputum transcriptomic response and NETosis indicating an adaptive immune response to *Haemophilus influenzae*. Therefore, a proportion of neutrophilic severe asthma may be driven by an overabundance of *Haemophilus influenzae* that may respond well to targeted antibiotic therapy.

Chapter 6. Host-microbial interactions differ with age of asthma onset

Parts of this chapter are based on the publication:

***A Versi**, A Azim et al. Host-microbial interactions differ with age of asthma onset. Eur Respir J 2024 Sep 5;64(3):2400428.*

6.1 Introduction

Asthma is characterised by chronic airway inflammation and is well recognized to be a heterogenous disease (90). Several distinct clinical phenotypes have been identified in asthma (275–277), the most common and easily recognized of which is allergic asthma. This phenotype is characterised by eosinophilic, or T2 airway inflammation, and often commences in childhood. Another recognized asthma phenotype is late-onset (LO) asthma, which refers to asthma that presents for the first time in adult life. This phenotype can be divided into T2-high and T2-low (270). Inflammation in late onset (LO) asthma also exhibits heterogeneity, particularly on the level of the relationship between airway neutrophilia and the airway microbiome (70,87,134,161).

There is a growing body of evidence supporting the association between airway microbiota in early life and the development of childhood asthma (75,76,271–280) as discussed in the introduction. The likelihood of asthma development by age 5 years was actually increased in infants who had pathogenic bacterial colonisation of the upper airways, in other words, abundances of certain pathogens such as *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Veillonella* and *Prevotella* species (271,273).

Describing asthma's heterogeneity by age of disease onset, variably dichotomised into early-onset (EO) and late-onset (LO) disease due to clear variations in patient characteristics, comorbidities and asthma control (281) also identifies differences in lung function decline (282) and response to anti-inflammatory therapies (269,283).

Taken together, these findings indicate that there are distinct mechanisms relating to airway inflammation that underlie EO and LO disease. Although limited, previous

efforts that have identified distinct severe asthma phenotypes based on sputum microbiome profiles have also recognized asthma onset and the duration of asthma as an important differentiating factor (70,87).

Moreover, the relationship between host-microbial interactions and age of asthma onset has not been fully established and several gaps persist in current knowledge. It is therefore critical to explore this relationship as it is increasingly recognised that age of asthma onset is important, not only for the diagnosis of the condition i.e. the differential diagnosis of suspected asthma varies with age (93) but also its management i.e. the response to asthma therapy varies with age of onset (284).

The hypothesis of this chapter was that host-microbial interactions in asthma may differ based on the age of asthma onset and these interaction differences may contribute to distinct inflammatory mechanisms.

In this analysis, we sought to apply a machine learning-based framework to describe differences in host-microbial interactions in the airways of EO and LO asthma. The use of this method addresses common difficulties in combining multi-omic data, such as high-dimensionality, sparsity, and multicollinearity, while also allowing for the discovery of existing interactions between airway microbes and host genes.

6.2 Method

6.2.1 *Participants*

The U-BIOPRED cohort consists of SA that included non-smokers and current and/or ex-smokers, and two control groups, MMA and HC as previously described in Datasets & Methodology (164). Participants provided induced sputum samples that passed quality control for metagenomic and transcriptomic analysis (20,170). All patients gave written informed consent to participate in the study.

6.2.2 *Induction and initial sample handling*

Sputum induction was undertaken and processed as previously described (20). Transcriptomic expression was performed using Affymetrix U133 Plus 2.0

(Affymetrix, Santa Clara, CA, USA) microarrays using RNA extracted from sputum cells. Frozen sputum samples were sent to Second Genome (San Francisco, California, USA) for metagenomic sequencing as previously described. The MetaPhlAn2 pipeline (*version 2.7.2*) and its marker database (10) was used to estimate microbiome profiles.

6.2.3 Statistical and computational analyses

Overall, we included 99 paired sputum metagenomic and transcriptomic samples from SA (n=79) that included non-smokers and current and/or ex-smokers, and MMA (n=20) subjects. Both the metagenomic and transcriptomic datasets have been previously published (20,170). These paired samples were split into two groups: EO and LO asthma using 18 years of age as the threshold. 43 samples were designated as EO and 56 as LO.

We employed a machine learning framework for integrating high-dimensional multi-omic data which is like a previously described approach (285).

Sputum transcriptomics pre-processing: Genes were filtered using a DEG analysis between SA (n=97) and HC (n=23) cohort using Limma (186) retaining genes which had an adjusted p-value < 0.05 and a $\log_2$ fold change (FC) > 0.5. Genes with low variance further filtered using a 25% quantile of variance across each, the LO group and EO group.

Sputum metagenomics pre-processing: Abundance based filtering was performed to retain species found at 0.01 relative abundance in at least 10% of samples across each group.

Sparse canonical correlation analysis (CCA) was used to identify paired host sputum gene expression and host sputum bacterial species abundance for each onset group. CCA identifies linear projection of two sets of observations into shared latent space that maximises the correlation between the two datasets. Sparse CCA is more suitable for high-dimensional datasets by utilizing lasso penalty for feature selection.

Objective function:

subject to

Sparse CCA was separately applied to each onset group using R 4.1.3 with the PMA package (version 1.2.1). The two variables to be correlated were the bacterial species relative abundance data from the sputum metagenomics and gene expression data from the sputum transcriptomics.

6.2.4 *Hyperparameter tunning*

Hyperparameter tuning was performed to identify parameters for sparsity penalties for both the sputum microbiome and gene expression datasets. A grid-search approach using leave-one-out cross validation was performed and the parameters which gave the highest correlation under cross-validation was chosen. Using this approach, the penalty parameters identified for EO were 0.15 and 0.45 and for LO were 0.10 and 0.52.

6.2.5 *Sparse CCA correlation*

After identifying the sparsity parameters, the sparse CCA model was fit to obtain subsets of correlated host genes and airway bacterial species, known as components. Each component includes non-zero weights (or canonical loadings) on the airway microbes and non-zero weights on a subset of host genes correlated with those airway microbes, to capture joint variation in the two datasets. The first 10 sparse CCA components were computed for each onset type and each component was computed iteratively, being informed from the previous component, so that all the components were uncorrelated.

6.2.6 *Pathway enrichment of genes*

To characterise host pathways enriched for the host sputum genes associated with sputum microbes in each component, a Fisher’s exact test was employed to assess pathway enrichment. The set of host genes that were input to the sparse CCA analysis was used as background genes, and the set of host genes in a component as the genes of interest were used to determine assess pathway enrichment. KEGG and Pathway Interaction Database (PID) gene sets from the MsigDB canonical

pathways collection (https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp) were used in the pathway enrichment analysis.

6.2.7 *Visualization of shared host pathways and disease-specific host pathways*

Shared host pathways are enriched host pathways for which host sputum gene expression correlates with airway microbes across the two onset groups. Disease-specific host pathways are enriched host pathways for which host sputum gene expression correlates with airway microbes for only one out of the two onset groups.

For visualizing shared host pathways across the two onset groups in addition to disease-specific host pathways for each onset group, pathways were selected (adjusted p-value <0.05) and the enrichment was visualised using ggplot2. For component visualisation corresponding to selected host pathways, both the host genes and species were ordered by their absolute weights in the components and the top 10 genes and species in terms of the weights were selected for representation. The sparse CCA components were visualised using ggplot2.

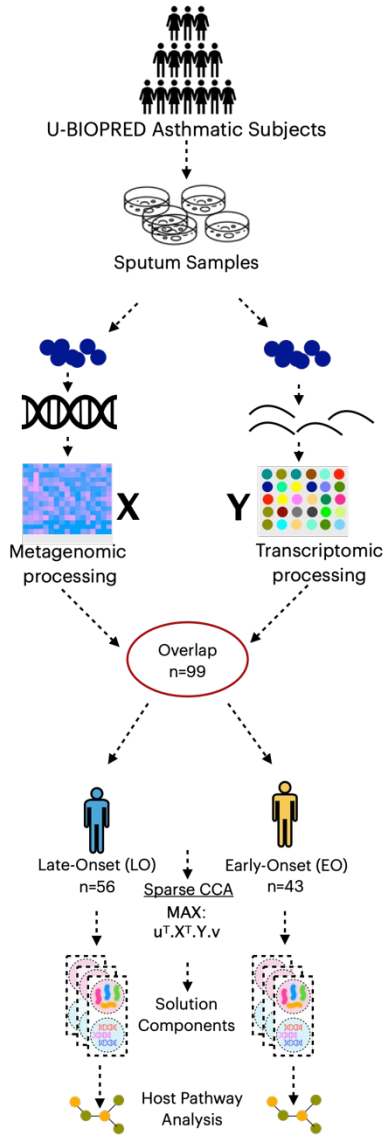
Three specific bacterial species, which are commonly implicated in severe asthma, were examined by checking whether the species is within the top species for each component ordered by absolute coefficient from sparse CCA. The corresponding pathways which were enriched by the genes of that component were selected for visualisation.

GSVA was used to calculate sample-wise enrichment score (ES) of a set of genes (gene signatures) in the normalised host sputum gene expression matrix. These signatures were derived from the components produced by sparse CCA, taking genes of the top 10 absolute weights for a component. The ES of these gene signatures were correlated with sputum neutrophils and sputum eosinophils using spearman’s correlation.

6.3 Results

The number of paired sputum transcriptomic and metagenomic samples (n=99) consisted of 79 SA and 20 MMA subjects. Out of the 99 paired samples, 56 samples were designated as LO and 43 as EO using a threshold of <= 18 years for the age of onset (**Figure 45**).

Figure 45. Flow Chart showing the overall method of sample collection, cohort grouping and the application of sparse CCA



Legend: Sputum samples were taken from U-BIOPRED asthmatic subjects (severe asthmatics (SA) and mild-moderate (MMA)). Metagenomic sequencing and transcriptomic

sequencing of samples was carried out. Grouping was done by age of asthma onset and sparse CCA was carried out. Pathway analysis of genes for each components were carried out.

There was a significant difference ($p<0.001$) in the duration of asthma between the LO (16.7±12.2 years) and EO (36.3±14.0 years) groups. LO patients had significantly higher blood (0.4±0.3 vs 0.3±0.2 $x10^3$ µL, $p=0.012$) and sputum eosinophils (15.8±21.1 vs 6.6±13.5 %, $p<0.05$) than EO patients (**Table 12**).

Table 12. Characteristics of Adult onset and Child onset

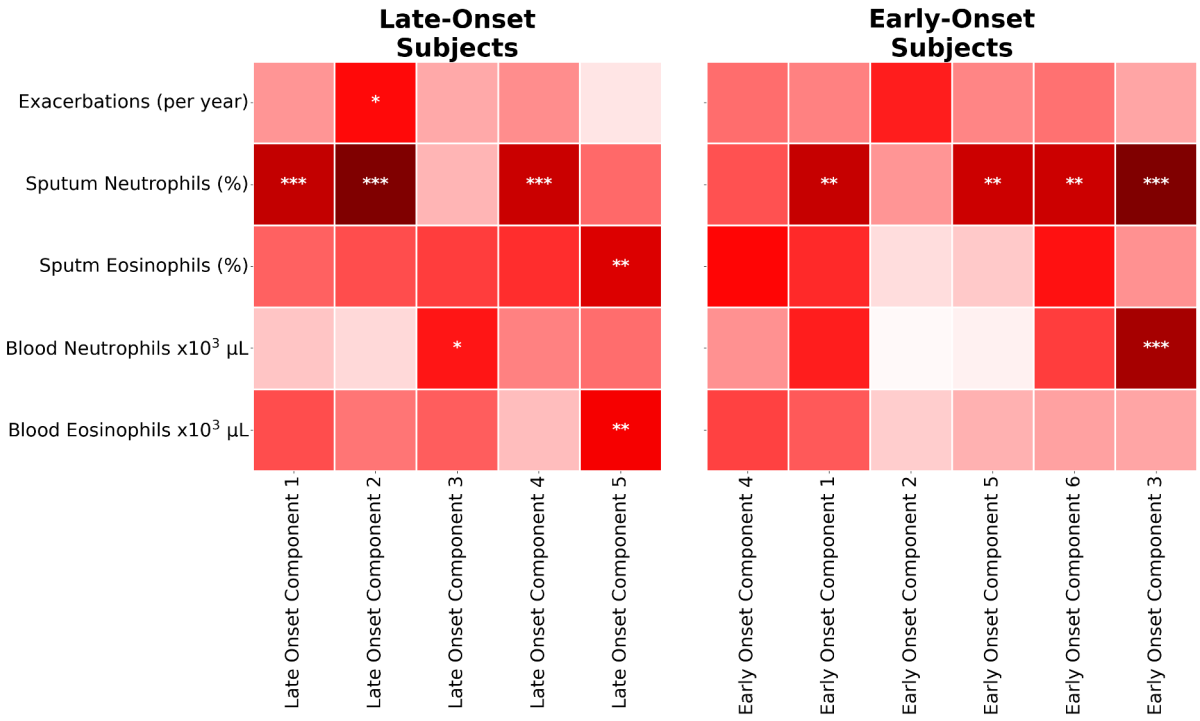
	Early-Onset	Late-Onset	#P value
Subjects, n	43	56	ND
Age (years)	46.0 [32.5,52.0]	57.5 [51.0,64.2]	NS
Females, n (%)	24 (55.8)	33 (58.9)	NS
BMI	26.7 (5.7)	28.6 (5.0)	NS
Duration (years)	36.3 (14.0)	16.7 (12.2)	<0.001
Current Smoking, n (%)	3 (7.0)	4 (7.1)	NS
Current & Ex-Smoking, n (%)	9 (20.9)	29 (51.8)	<0.05
Oral corticosteroid daily, n (%)	12 (27.9)	21 (37.5)	NS
ACQ-5 score	2.0 (1.4)	1.9 (1.2)	NS
Exacerbations, per year	1.8 (2.1)	2.1 (2.1)	NS
Nasal polyposis, n (%)	9 (20.9)	25 (44.6)	<0.05
Eczema, n (%)	17 (39.5)	13 (23.2)	NS
Allergic rhinitis, n (%)	24 (55.8)	17 (30.4)	<0.05
GERD, n (%)	16 (37.2)	22 (39.3)	NS
Blood Eosinophils μL^{-1}	0.3 (0.2)	0.4 (0.3)	<0.05
Blood Neutrophils μL^{-1}	5.0 (2.6)	5.2 (2.2)	NS
Sputum Neutrophils (%)	58.9 (26.1)	54.3 (25.3)	NS
Sputum Eosinophils (%)	6.6 (13.5)	15.8 (21.1)	0.01

Sputum Lymphocytes (%)	1.4 (1.3)	1.4 (1.2)	NS
Sputum Macrophage (%)	33.4 (24.6)	28.4 (21.7)	NS
FEV1 (% predicted)	71.4 (24.6)	68.7 (20.3)	NS
FeNO (ppb)	35.2 (34.2)	43.6 (38.3)	NS
Smoking pack years	12.3 (9.6)	13.5 (16.8)	NS
Antibiotic current, n (%)	6 (14.0)	10 (17.9)	NS
Serum CRP (mg/L)	4.7 (9.6)	6.6 (14.7)	NS
Serum IL-1α (pg/ml)	33.6 (7.5)	35.1 (7.5)	NS
Serum IL-6 (pg/ml)	1.0 (0.8)	1.3 (1.1)	NS
Serum IL-8 (pg/ml)	5.0 (8.2)	3.8 (1.6)	NS
Serum C5a (pg/ml)	39.4 (20.6)	45.0 (29.0)	NS

Abbreviations:
ACQ5: Asthma Control Questionnaire score; BMI: Body mass index; CRP: Serum C-Reactive Protein; FeNO: Fractional exhaled nitric oxide; FEV1: Forced Expiratory Volume in the first second pre-salbutamol; FVC: Forced Vital Capacity; GERD: Gastroesophageal reflux disease; N/A: Not applicable; NS: Not significant; ppb: parts per billion. Data shown as mean (Standard Deviation) unless variable is categorical where n (%). Age shown as median [95% Confidence intervals].
[#]P value: One-way ANOVA (continuous) & Chi-squared (categorical) tests P-value comparing the two cluster groups: LO & EO.

Compared to EO patients, LO subjects showed a significantly higher percentage of subjects with nasal polyposis (44.6% vs 20.9%, p=0.03) and had significantly higher blood (0.4±0.3 vs 0.3±0.2x109/L, p=0.012) and sputum eosinophils (15.8±21.1 vs 6.6±13.5%, p<0.05). In contrast, EO patients were more atopic (93.0% vs 58.9%, p=0.001), with a greater number of subjects having allergic rhinitis (55.8% vs 30.4%, p=0.03) and hay fever (65.1% vs 41.1%, p=0.02). However, no significant differences were observed between the two groups in terms of sputum neutrophils, sputum macrophages, current use of antibiotics, age, lung function (FEV1), FeNO, BMI, smoking status, oral corticosteroid use, ACQ5 scores, exacerbations, blood granulocyte levels, and serum biomarkers (**Table 12**).

Figure 46. GSVA sample-wise enrichment score correlation with clinical characteristics for each component



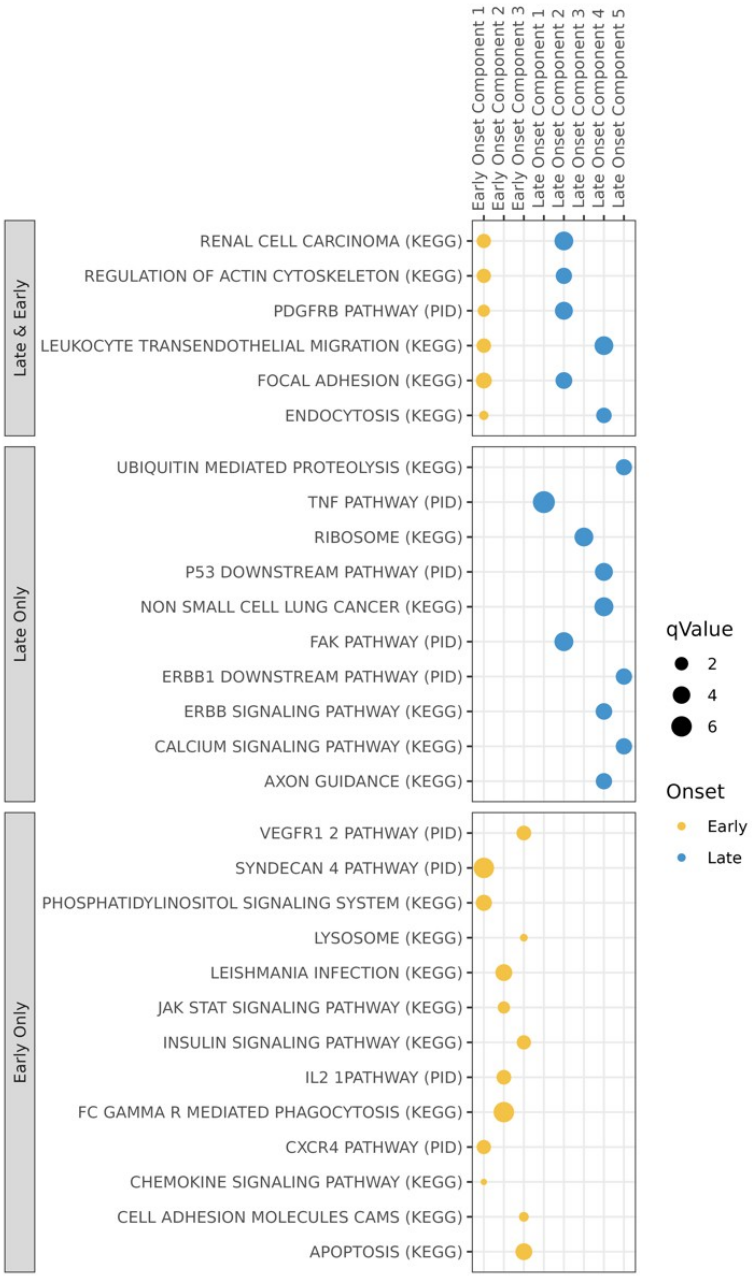
Legend: GSVA sample-wise enrichment score correlation with clinical characteristics showing the correlation of the ES of each gene set from all sparse-CCA derived components with sputum neutrophils and eosinophils.

The sparse-CCA analysis produced 5 components in LO asthma (LOC1-5) and 6 in EO asthma (EOC1-6). Three out of five components in LO asthma (LOC1, 2, 4) [(r=0.36, p<0.01), (r=0.70, p<0.001), (r=0.35, p<0.01) respectively] and four of six components in EO asthma (EOC1, 3, 5, 6) [(r=0.36, p<0.05), (r=0.63, p<0.001), (r=0.35, p<0.05), (r=0.35, p<0.05) respectively] had a gene set whose sample-wise ES using GSVA correlated with sputum neutrophilia (

) in the sputum gene expression matrix of their respective onset-group. One component (LOC5) had a gene set whose ES correlated with both sputum (r=0.32,

p<0.05) and blood eosinophilia (r=0.23, p<0.05) in the LO subjects and was dominated by *Neisseria* and *Haemophilus influenza.*

Figure 47. Component-genes pathway analysis



Legend: Component-genes pathway analysis showing the enrichment of pathways for each component using KEGG and PID pathway databases

Pathway enrichment analysis of the sparse-CCA produced components using the KEGG and PID databases revealed that all five components in LO asthma saw enrichment of pathways. However, only three out of six components in EO asthma had pathways enriched. Across all components pathway enrichment identified 6 pathways that were shared between the LO and EO phenotypes (**Figure 47**). These were renal cell carcinoma, regulation of actin cytoskeleton, leukocyte transendothelial migration, focal adhesion and endocytosis from the KEGG database and Platelet-derived growth factor receptor-beta signalling pathway from PID database. In addition, 10 pathways were only enriched in LO asthma and 13 pathways only enriched in EO asthma.

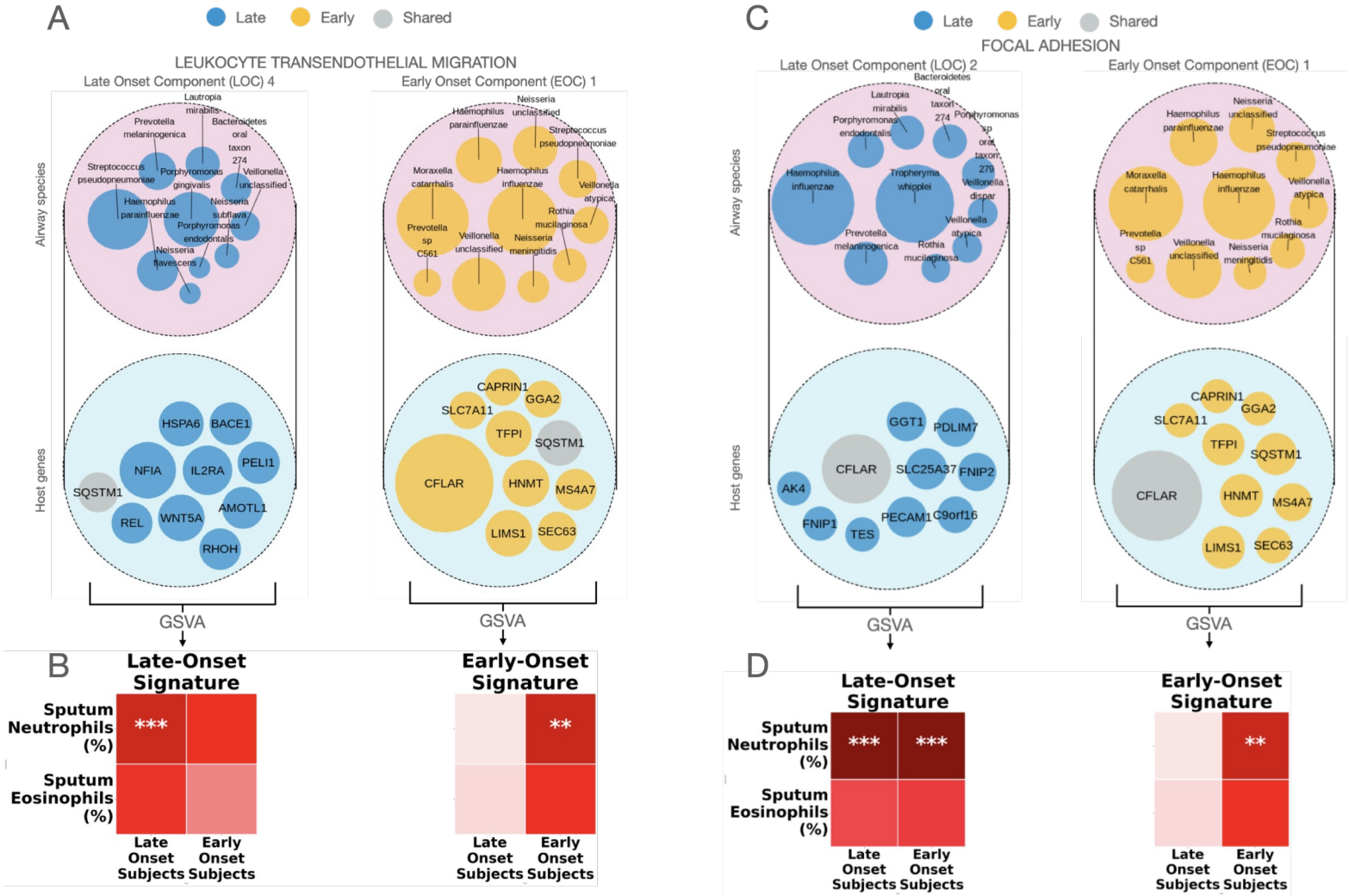
The shared pathways enriched in both LO and EO phenotypes, particularly involved components EOC1 and LOC2 and 4 (**Figure 47**). However, the components in question had different constituent genes and species, which were not identical. LOC4 and EOC1 were both enriched for leukocyte trans-endothelial migration, but LOC4 was characterised by *Streptococcus species* whilst EOC1 was characterised by *Moraxella catarrhalis* and *Haemophilus influenzae* with the shared gene of SQSTM1 in both components (**Figure 48A**). GSVA analysis revealed that the ES of genes in LOC4 were correlated with sputum neutrophilia in LO patients, but not in EO asthma. On the other hand, genes in EOC1 were correlated with sputum neutrophilia in EO asthma, but not in LO patients (**Figure 48B**).

Furthermore, LOC2 and EOC1 were both enriched for focal adhesion and both components were characterised by *Haemophilus influenzae* and the gene CFLAR. But EOC1 was again characterised by *Moraxella catarrhalis* whilst LOC2 was characterised by *Tropheryma whipplei* (**Figure 48C**). The GSVA analysis revealed ES correlation for LOC2 with sputum neutrophilia in LO patients and EO patients. However, EOC1 GSVA analysis ES only correlated with sputum neutrophilia in EO patients and not in LO patients.

Haemophilus influenzae was associated with pathways enriched in both EO and LO components (15 in EO and 10 in LO), meanwhile *Moraxella catarrhalis* was only associated with pathways enriched in EO (19 pathways) and *Tropheryma whipplei* with pathways enriched in LO (6 pathways) (**Figure 49**). LOC 3 had a set of genes,

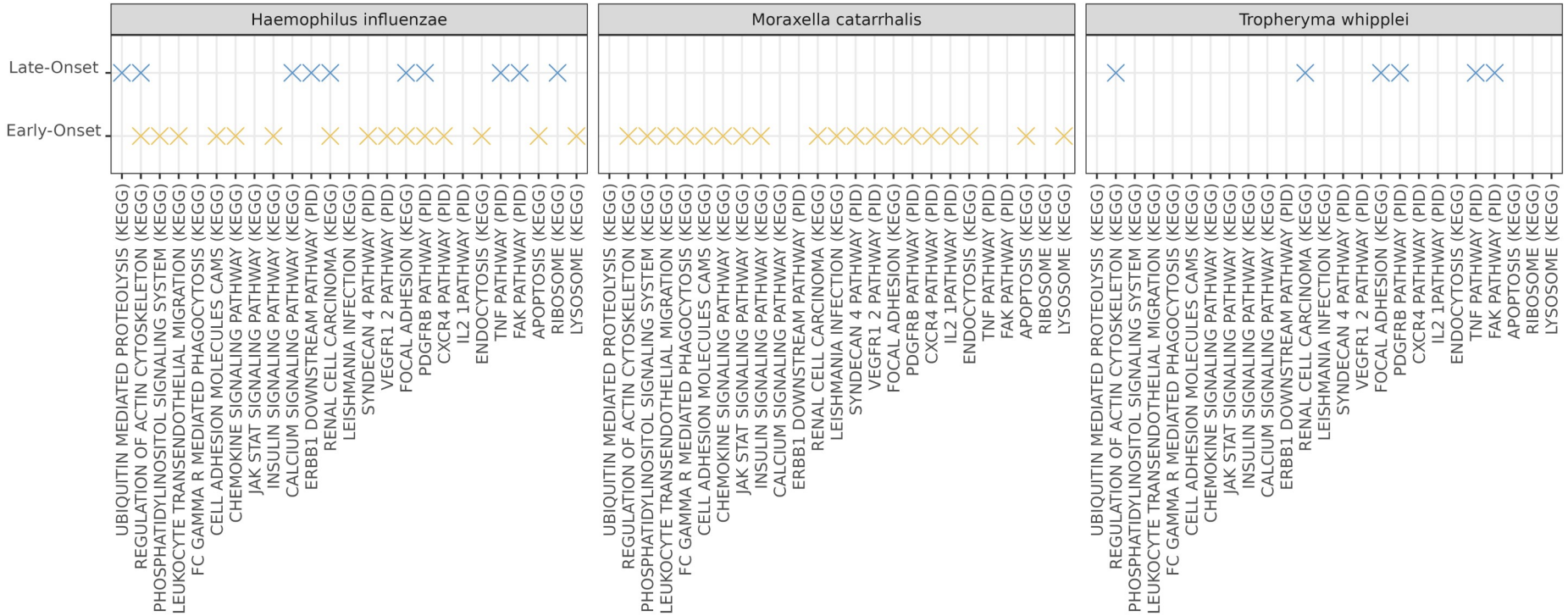
whose ES was correlated with sputum eosinophils ($r=0.32$, $p<0.05$) and was dominated by *Neisseria* and *Haemophilus influenza.*

Figure 48. Visualization of Leukocyte Transendothelial Migration and Focal Adhesion for both LO and EO asthma and GSVA sample-wise enrichment score correlation with clinical characteristics



Legend: Visualization of Leukocyte Transendothelial Migration and Focal Adhesion for both LO and EO asthma showing the top 10 genes and species by absolute weights and GSVA sample-wise enrichment score correlation with clinical characteristics showing the correlation of the ES of each gene set from the components in the Leukocyte Transendothelial Migration and focal adhesion pathway with sputum neutrophils and eosinophils.

Figure 49. Pathway analysis by dominant species in components for each onset-group



Legend: Pathway analysis by the species *Haemophilus influenzae, Moraxella catarrhalis* & *Tropheryma whipplei* for both onset groups LO and EO asthma showing the pathway enrichment of components for components which include *Haemophilus influenzae, Moraxella catarrhalis* & *Tropheryma whipplei* in the top 10 species by absolute weights.

6.4 Discussion

By applying sparse-CCA this analysis was able to demonstrate differences in EO and LO asthma. Using a machine learning-based multi-omic integration framework, common and phenotype-specific interactions between the airway gene expression and microbiome were identified. The utility of integrating the sputum microbiome and host gene expression together to obtain insight into their contribution to the disease process, is superior to single-dataset omics alone, as this approach can provide some information about the cross-talk of the complex network of interaction that occur in the airways which underlie the complex trait (286,287).

Pathway analysis of components associated with sputum neutrophilia in both EO and LO asthma identified upregulation of genes associated with transendothelial leukocyte migration. This broad observation is consistent with the established understanding that impaired barrier defences due to dysbiosis can lead to persistent airway neutrophilia (288).

This analysis also identifies upregulation of genes unique to each phenotype. In EO asthma, there is greater enrichment of genes associated with adhesion molecules. Compared to LO disease, there is a stronger transcriptomic contribution to EO asthma (289,290) particularly barrier function. This is corroborated by a greater association of atopic conditions in EO asthma (291). Furthermore, the analysis identified distinct correlations of LOCs and EOCs with sputum neutrophilia indicate that these host-microbial interactions are unique to each neutrophilic asthma phenotype.

Epithelial cell dysfunction is frequently observed in asthma. Loss of epithelial junctions and airway epithelial barrier function in asthma results from environmental, genetic and epigenetic factors and has been shown to increase host susceptibility to pathogens and allergens while also propagating pro-inflammatory responses in both EO and LO asthma (292). Bacterial airway colonisation is implicated in this pathway, as *S. pneumoniae* infection can cause epithelial barrier dysfunction (293) and the risk of asthma increases with bacterial colonisation of the airways. Given this

association we see in several components containing both *Haemohilus influenzae* and the gene CFLAR it stands to reason that a similar interaction with *Haemophilus*-CFLAR should be observed in both EO and LO asthma.

When exploring potential modifiers of the asthma phenotype, polymorphisms in CFLAR were found to be associated with asthma (294). CFLAR expression in the epithelium can be affected by environmental factors and was shown to be reduced by exposure to cigarette smoke both in vitro and in vivo (283). The downregulation of CFLAR expression results in damage-associated molecular pattern (DAMP) release, which in turn may increase susceptibility to immunogenic cell death. *Haemophilus influenzae* could provoke a similar process, as it is recognized that inflammasome activation in response to DAMPs (or pathogen-associated molecular pattern (PAMPs)) is one of the primary consequences of epithelial infection with non-typeable *Haemophilus* influenzae (269). This is corroborated by asthma mouse models showing that the inflammasome, including CASP8 (which is related to CFLAR), is associated with epithelial barrier function (295). Moreover, in vitro studies support the ability of bacteria to provoke CFLAR pathways in human epithelial cells (296,297).

Whilst the composition of the airway microbiome changes throughout life, it is particularly dynamic in the early years of life, when perturbations are thought to be critical to lower airway immune maturation (298). Evidence of this was observed in this analysis, where IL2 signalling is enriched in EO asthma but not in LO asthma. IL-2 is essential for regulatory T-cell function and survival, necessary components of immune homeostasis (299).

One limitation of this study is the use of absolute weights from the sparse CCA components. A disadvantage of taking absolute weights is that the direction of association is lost and therefore the gene set used for GSVA would include both upregulated genes and downregulated genes associated with the increase or decrease in abundance of the species in the respective component. Another limitation is that the LO asthmatic group contains significantly more current and ex-

smoking subjects than the EO group, which may confound the results with host-gene interaction based on smoking status rather than on onset-group alone.

In conclusion, analysis has revealed several host gene-microbiome associations and the combination of Wnt5a and CFLAR with SQSTM1, on microbial growth and on host-microbe immune interactions, which was not revealed in the previous chapter using a single-omic dataset approach rather than an integrative one. These associations may be a fruitful target for further studies. However, it is not possible to determine dynamic (time-dependent) conclusions regarding causality in this cross-sectional study. Furthermore, in this study microbial exposures cannot be accounted for and will clearly influence disease evolution and therefore a temporal microbiomics approach may be required to identify underlying endotypes (287). In addition, since U-BIOPRED is an adult cohort, the age of disease onset is confounded by the duration of the disease and treatment. Nevertheless, previous microbial analysis of U-BIOPRED data has revealed differences in microbial profiles and age of onset (87). This analysis expands on those findings by identifying shared and unique host-microbial interactions between EO and LO phenotypes. The results suggest that a better understanding of the heterogeneity of asthma immunopathophysiology can be achieved through the examination of host-microbial interactions.

Chapter 7. Summary & Discussion

The objective of this thesis was to gain insight into the potential of metagenomic data in enhancing the current understanding of the relationship between the airway microbiome and various phenotypes of asthma. Current approaches of the examination of the airway microbiome typically rely on 16S rRNA sequence approaches or target PCR, which have limitations in terms of species-level resolution (300). The usage of shotgun metagenomics data presents a new approach which overcomes this limitation and provides functional profiling of the microbiome, giving us understanding of the metabolism of the microbiome (213). Metagenomic profiling was carried out to better comprehend the intricate relationships that might exist between bacterial species, inflammatory pathways, and clinical parameters in asthmatics and healthy controls. Furthermore, it employed an additional omics platform, that of paired sputum transcriptomics, to get a more detailed understanding of the interactions that may occur between airway microbes and host gene expression. The utilization of both these datasets with well characterized clinical data presents a new prospect in the understanding of asthma.

7.1 Summary

7.1.1 Data processing and pipeline evaluation

First, the analysis in chapter 3 examined the outcome of metagenomic profiling using three different profiling tools: MetaPhlAn2, kraken and kaiju. MetaPhlAn2 emerged as the preferred choice compared to the other two programs due to its more realistic results. Another advantage lies in the high speed and low-memory requirements of MetaPhlAn2's efficient approach aligning sequencing reads to marker genes (207). Moreover, MetaPhlAn2 was able to address the limitation introduced by the U-BIOPRED samples that were used; the samples spanned the years 2010-2015, at which time the standard of read length was a 100bp paired-end sequencing protocol (shorter than the current standard of 250bp). By contrast, Kaiju demonstrates lower sensitivity for classifying reads with approximately 100bp in lengths compared to longer, 250bp reads (40% vs 70% sensitivity), particularly at the genus level (211).

This could account for the results obtained via Kaiju in our study, which were dominated by unexpected microbes including *Ralstonia*, *Klebsiella* and *Escherichia*. MetaPhlAn2 thus was established as a preferable program for microbiome profiling of the U-BIOPRED samples, despite some samples not being mapped to known reference microbial markers and, consequently, being excluded from further analysis. However, it is important to also acknowledge that MetaPhlAn2's performance can be limited by uneven distribution of sequence markers across microbes of interest, which introduces bias (207). Another limitation of note to the taxonomic classification and abundance data was not related to the program used for microbiome profiling, but rather the sequencing platform used to characterize the samples. The Illumina HiSeq sequencing platform has been associated with GC content bias, leading to an underrepresentation of organisms with low-or high-content GC regions (222–224). Finally, it is important to consider that we were unable to account for batch specific contaminants due to the lack of negative control samples for each batch. Despite this, we were able to establish that differences in the taxonomic abundance between batches were non-significant.

7.1.2 Microbiome evaluation by phenotype

In chapter 4, the analysis explored the abundance of bacterial species present in the sputum of severe asthmatics and how they may relate to asthma severity, inflammatory status, and molecular phenotype. Using metagenomic sequencing of samples from one of the largest cohorts of severe asthmatics, three pathogenic microbial species with potential pathogenicity in respiratory diseases were selected, namely *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Tropheryma whipplei*. Four severity cohorts were examined: SAn, SA/ex, MMA, and HC. Despite the reduction in bacterial α-diversity observed across study groups, significant differences in microbial species were evident between patient cohorts; *Haemophilus influenzae* and *Moraxella catarrhalis* were most abundant in the SAn cohort (vs. MMA, HC, and SA/ex), while the SA/ex group had a higher abundance of *Haemophilus influenzae* and *Tropheryma whipplei* compared to the MMA cohort (but not others). Reductions in several bacterial species were observed in our study in

severe asthmatics, irrespective of smoking status, including the previously reported asthma-depleted *Prevotella intermedia* (56,138).

Delving deeper into potential associations with inflammatory status, it was observed that neutrophilic inflammation and activation of neutrophils and inflammasome were associated with variation in the abundance of some bacterial species; α- diversity was the lowest in the mixed granulocytic group but only slightly reduced in the neutrophilic group. Of note that neutrophilic inflammation was also associated with increased abundance of *Haemophilus influenzae* and *Moraxella catarrhalis*, while elevated levels of *Tropheryma whipplei* heralded eosinophilic inflammation. *Moraxella catarrhalis* was positively correlated with sputum neutrophils in both current and ex-smokers with severe asthma (but not MMA or HC). *Prevotella* species were negatively correlated with sputum neutrophilia, and lower asthma exacerbations could be predicted by increased abundance of *Prevotella melaninogenica* and *Prevotella intermedia*.

The relationship between microbial species and molecular phenotypes in severe asthma was also examined. Notably, TAC1 and TAC2 were associated with the highest abundance of *Haemophilus influenzae*, which was intriguing given the presence of *Haemophilus influenzae* in both eosinophilic (TAC1) and neutrophilic (TAC2) phenotypes, but not in the paucigranulocytic phenotype (TAC3). While the increased abundance of *Haemophilus influenzae* in the sputum could be accounted for (reduced phagocytosis by lung macrophages in severe asthma) (228), the expected cascade triggering neutrophil activation and inflammation (229,230) is at odds with the higher presence of *Haemophilus influenzae* in both TAC1 and TAC2, and the lack of an evident association with oral corticosteroid therapy. By contrast, *Haemophilus parainfluenzae* was less present in non-smokers (SAn) as well as both TAC1 and TAC2. Moreover, our finding of reduced abundance of *Haemophilus parainfluenzae* in current smokers (vs. ex-smokers) calls into question the role of current smoking in reducing this bacterial species as this particular change in the bacterial microbiome could be due to OCS therapy.

At the time of writing, our findings concerning smoking status as a predictor of higher *Tropheryma whipplei* abundance and pack years of smoking as a (positive) predictor of eosinophilic (TAC1) and neutrophilic (TAC2) phenotypes remains the only evidence of the implication of smoking on sputum metagenome of severe asthmatics. The mechanism and potential pathogenesis of *Tropheryma whipplei* in smoking/ex-smoking severe eosinophilic asthma remains uncertain. As for, *Moraxella catarrhalis*, its higher abundance in TAC2 combined with its association with sputum neutrophils supports its link with neutrophilic airway inflammation reported in pathogenic respiratory infections (91,134). Whether *Morexella catarrahlis* drives sputum neutrophilia in these patients or whether other factors are driving the presence of both awaits targeted investigation. A vaccine against non-typeable Haemophilus influenza and Moraxella catarrhalis is safe in patients with COPD although it did not significantly prevent exacerbations in proof-of-concept phase 2b study (301). Furthermore, a time course analysis of azithromycin, a macrolide that preferentially targets gram-negative bacteria such as *Haemophilus influenzae* and prevents exacerbations in patients with neutrophilic asthma (241), may determine whether any effect on *Haemophilus influenzae* precedes sputum neutrophilia.

Long-term follow-up data indicated relative stability in the microbiome of severe asthma patients over 12 months in two distinct robust severe asthma phenotypes (87), with the caveat that while the microbiome may remain stable, disease worsening and exacerbations can still occur, engendering microbiome changes (239). Worsening asthma outcomes, increased sputum neutrophilia, and higher abundance of certain bacterial species were identified in the two identified severe asthma phenotypes, with the possible implication of a genetic component (240). A limitation of the earlier study by Abdel-Aziz and co-workers (87) is that the incorporation of transcriptomic analysis was undertaken after the identification of 2 metagenomic clusters rather than utilising a truly integrated analysis.

The technical limitations of the analysis of the analysis were acknowledged, including the choice of MetaPhlAn2 (potential decrease in taxonomy profiling accuracy), the choice of algorithmic marker-based approach for taxonomic assignment (potential incomplete results) and incomplete data on medication history.

Despite these limitations, the findings contribute to understanding the complex interplay between microbial species, inflammatory pathways, and disease severity in asthma, emphasizing the need for further research to confirm and explore the implications of specific bacterial species in asthma pathogenesis.

7.1.3 Relative dominant species based phenotypes

The analysis from chapter 5 aimed to quantify the frequency of relative dominant bacterial species in the sputum of severe asthma patients, introducing a novel approach based on linking results from severe asthma patients to a state of significantly low diversity variance observed in control groups (mild-moderate asthma and healthy subjects). This approach accounts for the major abundance shifts that lowly-abundant taxa are prone to (256,257), allowing the identification of sputum samples with a high relative abundance of a particular species, termed Rare Dominant Species (RDS). Based on this approach, we were able, for the first time, to demonstrate that a restricted phenotype of high relative abundance of *Haemophilus influenzae* in neutrophilic severe asthma is associated with an adaptive response of NETosis and Th17 activation. The results require replication and testing in *in vivo* models to show causality.

Low α-diversity in severe asthma was associated with the relative dominance of 25 bacterial species (many of unknown pathogenic significance), with *Haemophilus influenzae* being the most prevalent (44%) RDS with known respiratory pathogenic effects, followed by *Moraxella catarrhalis* (14.4%) and *Tropheryma whipplei* (14.4%).

The clinical characteristics showed, longer asthma duration, increased exacerbations, high use of oral corticosteroids, sputum neutrophilia, lower sputum macrophages and sputum eosinophilia were associated with *Haemophilus influenzae* RDS. *Moraxella catarrhalis* RDS was solely associated with high sputum neutrophils but low sputum eosinophils and better outcomes (FeNO, asthma control), while *Tropheryma whipplei* RDS correlated with reduced FEV1, indicative of severe airflow obstruction and type 2 inflammation.

Clustering of bacterial RDSs samples produced a cluster with high relative abundance *of Haemophilus influenzae*. While small, this cluster of dominant Haemophilus influenzae RDSs termed C2-Hi was associated with long asthma duration, neutrophilic inflammation (high sputum neutrophils, low sputum eosinophil and macrophage), and low T2 inflammation. More importantly, C2-Hi was linked, for the first time, to activation of pathways related to MAPK, NF-κB, TNF, mTOR, and cellular necroptosis, as well as to a high expression of neutrophil extracellular traps (NETs). These are web-like scaffolds of DNA complexing with histones and neutrophil granular proteins released from neutrophils upon activation by pro-inflammatory stimuli (256). The high expression of the gene signature for NETs in C2-Hi suggests a host-bacterial response to high relative abundance (>45%) of *Haemophilus influenzae* in the airways, potentially contributing to NETopathic inflammation and airway epithelial damage (260,302).

Moraxella catarrhalis RDSs were associated with pathways enriched for the innate immune system consistently with that of *Haemophilus influenzae* RDSs, with additional activation of granulocyte, and macrophage colony-stimulating factors. This confirmed previous reports in the literature (265,266). By contrast, *Tropheryma whipplei* RDSs showed no association with thee immune-activated pathways.

It is important to note that while effective, the analysis' approach of characterizing clusters based on species' relative abundance led to the loss of other species, including *Rothia mucilaginosa*, which may have anti-inflammatory (267) or bioprotective properties in asthma (268). Despite the other limitations, such as the relatively low number of samples constituting RDS and the compositional nature of datasets generated for the metagenomic sequencing (only relative species abundance available), the study offers crucial insights; It is the first to delineate a phenotype characterized by a high relative abundance of *Haemophilus influenzae* in neutrophilic severe asthma, associated with an adaptive response of NETosis and Th17 activation. The clinical implications suggest that patients with this cluster may potentially respond well to targeted antibiotic therapy, such as macrolides, to which *Haemophilus influenzae* is usually sensitive (241). Further research with larger

cohorts and longitudinal studies is recommended to confirm and explore these findings.

7.1.4 Microbiome-host gene expression interaction analysis

Following on from the identification and analysis of RDSs, one interesting finding of this analysis is the distinct difference in the age of onset and duration of disease between RDS and non-RDS samples as well as the *Haemophilus influenzae* rich cluster (C2-Hi) vs the non-RDS group. Existing evidence have described the impact of airway microbiota in early life on the development of EO asthma (271,75,272–277,76,278–280). However, very few if any studies have explored the association between the lung microbiome and age of disease onset in adult populations despite the clear differences in clinical outcomes (asthma control, lung function decline and response to anti-inflammatory therapies) between EO and LO asthma (281,282,284). A distinct phenotype has been linked with early-age of asthma onset, be it on the level of the lung microbiome (70,87), or gut microbiota (303). To understand this phenotype better and address existing gaps in the literature, a multi-modal machine learning-based framework (285) to describe differences in host-microbial interactions in the airways between EO and LO asthma was undertaken.

The pathway enrichment of the components produced by Sparse CCA produced 6 pathways which were shared between both onset-groups. However, the genes and species which constitute these components were not identical. EO asthmatics' components where characterised by *Moraxella catarrhalis* and *Haemophilus influenzae*. Both onset-groups have a component which was enriched for leukocyte transendothelial migration. GSVA of genes in the LO group, but not in patients with EO asthma, correlate with sputum. Conversely, genes in EO component, but not LO asthma, correlate with sputum neutrophilia.

7.2 Discussion

7.2.1 Impact on Landscape

This analysis is the first to investigate the airways of asthmatics with paired sputum metagenomics and transcriptomics. Shotgun metagenomic sequencing from sputum induction had previously only been conducted on a small number of subjects (n=55) (304) at the start of this thesis. Research has revealed the link between smoking and altered microbial diversities and communities in the lungs (84,305). However, the effect of smoking on the lung microbiome of the upper respiratory tract is variable; in one study, there was an increased abundance of *Megasphaera*, *Veillonella* spp. and depletion of *Capnocytophaga*, *Fusobacterium*, *Neisseria* spp. in the oropharynx of healthy (non-asthmatic) smokers, while *Eggerthella, Erysipelotrichaceae* I.S., *Dorea*, *Anaerovorax* and *Eubacterium* spp. were enriched in the nasopharynx and *Shigella* spp. was depleted (305). In another study, smoking in the general population was associated with significant disruptions to the airway ecosystem characterized by the lower abundance of *Neisseria* spp. and the increase in *Streptococcus* spp. in posterior oropharyngeal swabs (84). It is possible that smokers might become more likely to develop recurrent infections due to this higher prevalence of *S. pneumoniae/mitis* clades, which may in turn lead to respiratory conditions such as COPD (305,306). The data also suggests that the site of sample collection and disease status of subjects may impact on the results.

The study from chapter 4 demonstrated the influence of smoking status on the sputum metagenome of severe asthmatics. Smoking predicted lower abundance of *Haemophilus parainfluenzae*. Consistently, the abundance of *H. parainfluenzae* and other bacteria such as *N. elongate* and *N. sicca* were found to be decreased in the lung of smokers (307). Another study also reported an increased relative abundance of *Haemophilus* as well as different genera (*Rothia, Neisseria, Actinomycetes, Porphyrins, Streptococcus*, and *Acinetobacter)* among smokers compared to non-smokers (308,309). By contrast, increased abundance of *Proteobacteria* has been described in asthma, including many known potential pathogens from the genera *Haemophilus*, *Moraxella*, and *Neisseria* (56,83–85,310). These pathogens likely contribute to asthmatic airway inflammation by causing intermittent mucosal damage.

The study also found higher *Tropheryma whipplei* abundance among smokers and ex-smokers. The mechanism and potential pathogenesis of *Tropheryma whipplei* in smoking/ex-smoking severe eosinophilic remains uncertain. Studies suggest a potential link between corticosteroid-resistant asthma and enrichment of certain genera, such as *Haemophilus* and *Tropheryma (311).* On the other hand, relative ICS insensitivity in asthma has also been linked to smoking (312,313). It is therefore possible that increased abundance of *Tropheryma whipplei and Haemophilus* could contribute to the development of a severe refractory asthma phenotype, both in smokers and non-smoking asthmatics.

Regarding asthma patients who are smokers, alteration of airway eosinophil and neutrophil numbers have been observed. This could explain the positive association found in the Chapter 4 study between pack years of smoking and both eosinophilic (TAC1) and neutrophilic (TAC2) phenotypes. However, evidence is variable, with some studies reporting reduced airway eosinophils (314–317), while other studies report unaltered eosinophils among smokers (318–322). A ≥10-pack-year history has been associated with increased eosinophilic airway inflammation in former smokers, in addition to autoimmunity toward eosinophils, and reduced sputum eosinophil sensitivity to systemic corticosteroids, indicating the effect of cumulative exposure to cigarette smoke on the emergence of a phenotype of severe refractory eosinophilic asthma (322). Similarly, the impact of current smoking on neutrophilic airway inflammation was inconsistent. Some data showed comparable airway neutrophil numbers between current smokers and never smokers (314,316,317,323), while other studies were able to link current smoking with neutrophilic airway inflammation (315,318,321).

As for asthma patients in general, smoking is not the only factor affecting asthma severity or airway inflammation, which are altered in response to variation in airway microbiota. The analysis from Chapter 4 found significantly reduced α-diversity among severe asthmatics compared to mild-moderate asthma patients and healthy controls, irrespective of smoking status. This is consistent with reports of lower α-diversity in asthmatics with evidence of airway obstruction (138). Reduced abundance of commensal bacteria has been described along with reduced α-

diversity, most likely leading to the growth of pathogenic bacteria (83,87). Enrichment of Proteobacteria (such as *Haemophilus influenzae*, *Moraxella catarrhalis* and *Streptococcus pseudopneumoniae)* has actually been linked to greater bronchial hyper-responsiveness and worse lung function (83,87). Consistently, our third study showed that patients with *Haemophilus influenzae* RDS had the longest duration of disease with evidence of more severe disease with highest number of exacerbations in the previous year and of patients on OCS therapy. This provides further support to the association between increase in abundance of pathogenic bacteria (such as *Haemophilus influenzae*) in the lung microbiome and the emergence of more severe asthma phenotypes.

Asthma endotypes also exhibit associations with airway microbiota. Lower bronchial bacterial burden, including a lower prevalence of pathogenic organisms in sputum (such as *Haemophilus*) has been associated with eosinophilic inflammation (113,135). This could possibly be due to the bactericidal activity of airway eosinophils in T2-high inflammation (324). In fact, enrichment of *Haemophilus, Moraxella, Neisseria Actinomyces* and *Rothia* has been described in non-eosinophilic or low eosinophilic asthma groups (85,90), while high eosinophilic groups had higher α-diversity and enrichment of other bacteria such as the genera *Halomonas, Aeribacillus* and the phylum Firmicutes (mainly *Streptococcus and Actinomycetaceae*) (70,90).

By contrast, the analysis from chapter 4 found a higher abundance of both *Haemophilus influenzae* and *Tropheryma whipplei* in eosinophilic inflammation. In addition, the eosinophilic (TAC1) asthma phenotype was found to have a higher prevalence of *Haemophilus influenzae*. While *Tropheryma whipplei* has been previously detected in sputum samples of severe asthma patients who were mainly eosinophilic (134).

Neutrophilic activation mediated by the inflammasome has been associated with increased abundance of several *Proteobacteria* including *Moraxella* and *Haemophilus*, and in particular *Haemophilus influenzae (134,161,242).* The analysis from chapter 4 found a higher prevalence of *Haemophilus influenzae* in the

neutrophilic (TAC2) asthma phenotype. This is consistent with the known activation of TLR2 by this pathogen and the subsequent cascade of inflammatory mediators (IL1α, IL1β, IL-6, IL-8, MCP-1 and TNFα) leading to neutrophil activation and inflammation in the epithelium (229,230). In chapter 5, a putative novel phenotype of severe asthma that is driven by a bacterial-host interaction, as evidenced by a *Haemophilus influenzae cluster* of the highest abundance associated with activation of neutrophilic pathways and NETosis, is described. This cluster of highly dominant *Haemophilus influenzae* being associated with sputum neutrophilia and evidence of Th-17 activation is consistent with the reported ability of *Haemophilus influenzae* to switch T2-mediated eosinophilic inflammation to a Th17-driven neutrophilic inflammation (261)

The new reincarnation of the hygiene hypothesis by Strachan in 1989, the microflora hypothesis, suggests microbial dysbiosis in early life may result in increased occurrences of allergy disorders (325). The early infant airway microbiome of those who are likely to develop EO asthma is characterised by colonisation of pathogenic bacteria *Moraxella, Haemophilus, Streptococcus, Veillonella* and *Prevotella* species (271,272). The analysis from chapter 6 demonstrates that EO asthma continues to have this association with *Moraxella catarrhalis* and is associated with the enrichment of leukocyte transendothelial migration and focal adhesion with its impact being neutrophilic in nature. This raises questions regarding the continued prevalence of *Moraxella catarrhalis* throughout the lifespan of EO asthmatics and how this develops and is linked to an increased propensity for reoccurring viral infections and its impact on the morphology of the airways.

Previous studies of the airway microbiome have found suggestions of an association with age of asthma onset. Clustering of U-BIOPRED patients with moderate to severe asthma based on clinicophysiologic parameters allowed the detection of a cluster (T2) with LO severe asthma with a history of smoking and chronic airflow obstruction (326). Later, microbiome analysis of U-BIOPRED confirmed the presence of distinct clusters relating to age of onset, identifying differences in microbial profiles and age of onset (87); eosinophilic inflammation, with implication of mast cells, group 3 innate lymphoid cells and ILC-3 pathways were associated with

LO severe asthma, while EO was linked to enrichment of genes related to lung injury and/or fibrosis (327). The analysis in Chapter 6 extends those findings by identifying shared and unique host-microbial interactions between EO and LO phenotypes.

The analysis established an association between EO asthma and *Moraxella catarrhalis* as well as neutrophilic inflammation. This is consistent with results from chapter 5, as well as the known propensity of *Moraxella catarrhalis* to cause respiratory infections mediated by neutrophilic inflammation in patients with severe or poorly controlled asthma (91,134). Infection with *Moraxella catarrhalis* in early life has been linked to a higher risk of asthma development of exacerbation of chronic pulmonary disease symptoms (75). Consistently, the analysis from chapter 6 also showed that EO asthma is associated with the enrichment of leukocyte transendothelial migration and focal adhesion. While upregulation of genes associated with transendothelial leukocyte migration were also observed in LO asthma, it is consistent with the impairment of barrier defences that is expected when dysbiosis occurs and that can result in persistent airway neutrophilia (288). As for the enrichment of genes associated with adhesion molecules in EO asthma, it is evident that epithelial barrier dysfunction can be caused by bacterial airway colonisation (293). When specifically considering *Moraxella catarrhalis,* this bacteria targets airway epithelial cells, leading to an increase in the expression of adhesion molecules which in turn induce the production of chemoattractant proteins to recruit monocytes and dendritic cells (328). The inflammatory cascade that is triggered by *Moraxella catarrhalis* ultimately results in the release of elastase by activated neutrophils, causing tissue damage (328). This complex interaction between *Moraxella catarrhalis,* neutrophilic inflammation leukocyte transendothelial migration and focal adhesion in EO asthma could explain the previously reported association of childhood-onset asthma and enrichment of genes related to lung injury and/or fibrosis in the U-BIOPRED study (327).

While the influence of *Moraxella catarrhalis* seemed to be restricted to EO asthma, *Haemophilus influenzae* was associated with both LO and EO groups. Seeing as bacterial colonisation of the airways increases the risk of asthma, it is possible that the interaction of *Haemophilus*-CFLAR should be observed in both EO and LO

asthma. Polymorphisms in CFLAR have actually been linked to asthma (294), and environmental factors such as exposure to smoking can reduce CFLAR expression (283). Studies have previously demonstrated that bacterial infections can provoke CFLAR pathways in human epithelial cells (296,297). It is therefore possible that epithelial infection with *Haemophilus influenzae* may decrease CFLAR expression through the release of DAMPs (or PAMPs) and the activation of the inflammasome that is characteristic of this pathogen (269).

7.2.2 Limitations

There are several limitations to the analyses described in this thesis, which are discussed in each relevant results chapter. There are two significant limitations that are not discussed previously but have an overarching impact on the thesis. Firstly, the lack of any additional dataset to validate the findings of each of the results chapters. Currently U-BIOPRED is unique in that it has paired sputum metagenomics and transcriptomics in well characterised subjects with ranging severity and healthy controls and therefore there exists no such other datasets with these characteristics. Although much of the findings are consistent with existing literature, the utilization of a validation cohort would give much more credence to the findings in the thesis and would allow us to understand to what extent the inherent limitation in metagenomic data of U-BIOPRED has had on the findings in this thesis.

The second fundamental limitation of the thesis is that the datasets are cross-sectional in nature and hence limits the outcomes of the analyses to associations alone, rather than any causal effect. However, these types of datasets allow for exploratory analysis and hypothesis building to narrow the field for further studies in man or in relevant disease/infection models.

Limitations associated with the metagenomic data are addressed in chapter 3. Due to the short read length of 100bp of the metagenomic data in U-BIOPRED, taxonomic classification of bacterial species may be prone to error. In addition, another limitation was the lack of sequencing of blank controls, which allows for the control of any contaminants in the reagents used.

Limitations of small number also impacted this thesis. As the dissection of prospective phenotypes occurred the number of observations decreased**.** This precipitated in the low numbers of subjects in the C1-Hi and C2-Hi *Haemophilus influenzae* RDS clusters in chapter 5. Although these clusters provide an interesting observation in severe asthma, the limited number of subjects, reduced the statistical power in any subsequent analysis of these clusters.

Clustering techniques also present some limitations, as these approaches aim to group subjects with similar profiles together and separate those with more distinct profiles. However, the choice of clustering technique and different preprocessing of the input data can have a direct impact on the outcome of clusters and how relevant they are (329). The use of bootstrapping was employed in chapter 5 to assess stability of the clusters by resampling, randomly removing 10% of the data and repeating the clustering through 1,000 iterations. Using this approach gives more confidence that the clusters produced were not artificially split, but rather due to inherit similarity in the metagenomic profiling.

7.2.3 Future Studies

The biggest limitation of these analyses is the lack of a validation dataset. If time and resources were permitting, conducting a study to collect sputum samples with paired metagenomics (with 250bp read-length) and transcriptomics (RNA-seq) derived (with blank controls sequenced) from a larger number of well characterised asthmatics and healthy controls would allow for the validation of these findings. The use of RNA-seq analysis of sputum pellets and of airway epithelial cell brushings, would allow for not only investigation into host gene expression, but also into meta-transcriptomics (looking at gene expression of the microbiome) in addition to RNA viruses. This dataset would then allow us to test our hypothesis on C2-Hi cluster in chapter 5 about whether there is a change to the glycolytic pathway for energy metabolism in this *Haemophilus* rich cluster, due to meta-transcriptomic data being available with the functional metagenomic data.

These datasets can then be used for the construction of a pan-genome scale metabolic model (pan-GEMs) (330,331), for both C1-Hi and C2-Hi clusters. This will

allow us to understand the distinctions between the two microbiome clusters in terms of flux through metabolic pathways and allows us to model the effects of different substrates on metabolic flux. In addition, we can also model the effects of introducing new species into these pan-GEMs to find species which will modulate the pathogenic activity of the C1-Hi and C2-Hi clusters. For example, does introducing species which are considered bioprotective, such as *Rothia mucilaginosa* or *Prevotella* species, modulate the pathogenic activity of the C2-Hi cluster. Furthermore, recent work has adapted GEMs to estimate cooperative and competitive indices for pairwise interactions between species (332). This type of model can be used to infer competitive species to those dominant pathogenic species.

Following validation in this new dataset, with the additional findings of the meta-transcriptomic profile of the airway microbiome, the question of causality can begin to be addressed with wet-lab experiments and using *in vivo* mice models allows for the assessment of environmental influence and long term effects (333). Although these models do have their weakness (333,334), they are considered a good step in determining causality in asthma (335). The usage of murine microbiome models have been established for looking at the lung microbiome (333,336). *Haemophilus influenzae* has previously been shown to promote asthma progression and induce Th17 immune responses after 16 days of exposure to non-typeable *Haemophilus influenzae* (261). Future work on the C2-Hi microbiome interaction on epithelial barrier function using murine microbiome models would allow for the testing of hypothesises gathered from chapter 5 and 6. Furthermore the addition of competitive species to these microbiomes using the results from the GEMs would allow us to test if the introduction of these species reduce the pathogenic burden of the microbiome.

Chapter 8. References

1. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M, et al. Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J. 2008 Jan 1;31(1):143–78.

2. Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee Report. Allergy. 2004;59(5):469–78.

3. Redd SC. Asthma in the United States: burden and current theories. Environ Health Perspect. 2002 Aug;110(Suppl 4):557–60.

4. Akinbami LJ, Simon AE, Rossen LM. Changing Trends in Asthma Prevalence Among Children. Pediatrics. 2016 Jan;137(1):1–7.

5. Asher MI, Montefort S, Björkstén B, Lai CKW, Strachan DP, Weiland SK, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. Lancet. 2006 Aug 26;368(9537):733–43.

6. Janson C, Anto J, Burney P, Chinn S, de Marco R, Heinrich J, et al. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. Eur Respir J. 2001 Sep;18(3):598–611.

7. Bloom CI, Saglani S, Feary J, Jarvis D, Quint JK. Changing prevalence of current asthma and inhaled corticosteroid treatment in the UK: population-based cohort 2006-2016. Eur Respir J. 2019 Apr;53(4):1802130.

8. Mukherjee M, Stoddart A, Gupta RP, Nwaru BI, Farr A, Heaven M, et al. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. BMC Med. 2016 Dec;14(1):1–15.

9. Nunes C, Pereira AM, Morais-Almeida M. Asthma costs and social impact. Asthma Res Pract. 2017 Jan 6;3:1.

10. Hekking PPW, Wener RR, Amelink M, Zwinderman AH, Bouvy ML, Bel EH. The prevalence of severe refractory asthma. J Allergy Clin Immunol. 2015 Apr;135(4):896–902.

11. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J. 2014 Feb;43(2):343–73.

12. Gross NJ. What is this thing called love? --or, defining asthma. Am Rev Respir Dis. 1980 Feb;121(2):203–4.

13. Fahy JV. Type 2 inflammation in asthma--present in most, absent in many. Nat Rev Immunol. 2015 Jan;15(1):57–65.

14. Spycher BD, Silverman M, Kuehni CE. Phenotypes of childhood asthma: are they real? Clin Exp Allergy. 2010 Aug;40(8):1130–41.

15. Chung KF, Adcock IM. Precision medicine for the discovery of treatable mechanisms in severe asthma. Allergy. 2019 Sep;74(9):1649–59.

16. Buhl R, Humbert M, Bjermer L, Chanez P, Heaney LG, Pavord I, et al. Severe eosinophilic asthma: a roadmap to consensus. Eur Respir J. 2017 May;49(5):1700634.

17. Gould HJ, Sutton BJ. IgE in allergy and asthma today. Nat Rev Immunol. 2008 Mar;8(3):205–17.

18. Banafea GH, Bakhashab S, Alshaibi HF, Natesan Pushparaj P, Rasool M. The role of human mast cells in allergy and asthma. Bioengineered. 2022 Mar;13(3):7049–64.

19. Kuruvilla ME, Lee FEH, Lee GB. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. Clin Rev Allergy Immunol. 2019 Apr;56(2):219–33.

20. Kuo CHS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. Eur Respir J. 2017 Feb 1;49(2):1602135.

21. Simpson JL, Phipps S, Baines KJ, Oreo KM, Gunawardhana L, Gibson PG. Elevated expression of the NLRP3 inflammasome in neutrophilic asthma. Eur Respir J. 2014 Apr;43(4):1067–76.

22. Wang Z, Li Y, Gao Y, Fu Y, Lin J, Lei X, et al. Global, regional, and national burden of asthma and its attributable risk factors from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. Respir Res. 2023;24:169.

23. Momtazmanesh S, Moghaddam SS, Ghamari SH, Rad EM, Rezaei N, Shobeiri P, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. eClinicalMedicine. 2023;59:101936.

24. Mortimer K, Reddel HK, Pitrez PM, Bateman ED. Asthma management in low and middle income countries: case for change. Eur Respir J. 2022 Sep;60(3):2103179.

25. Scholes S, Mindell JS. Health Survey for England 2018 Asthma [Internet]. 2019 [cited 2023 Sep 18]. Available from: http://healthsurvey.hscic.gov.uk/media/81643/HSE18-Asthma-rep.pdf

26. NHS warning to parents as ‘asthma season’ hits [Internet]. 2019 [cited 2023 Sep 18]. Available from: https://www.england.nhs.uk/2019/09/nhs-warning-to-parents-as-asthma-season-hits/

27. Gupta RP, Mukherjee M, Sheikh A, Strachan DP. Persistent variations in national asthma mortality, hospital admissions and prevalence by socioeconomic status and region in England. Thorax. 2018 Aug;73(8):706–12.

28. Gokhale M, Hattori T, Evitt L, Lenney W, Nordstrom B, Collins J, et al. Burden of asthma exacerbations and health care utilization in pediatric patients with asthma in the US and England. Immun Inflamm Dis. 2020 Jun;8(2):236–45.

29. NHS Long Term Plan: Respiratory disease [Internet]. 2023. Available from: https://www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/better-care-for-major-health-conditions/respiratory-disease/

30. Murdoch JR, Lloyd CM. Chronic inflammation and asthma. Mutat Res. 2010 Aug 7;690(1–2):24–39.

31. Lloyd CM, Robinson DS, McMillan SJ. Modelling airway remodelling in asthma. Drug Discovery Today: Disease Models. 2004 Dec 1;1(4):425–30.

32. Gans MD, Gavrilova T. Understanding the immunology of asthma: Pathophysiology, biomarkers, and treatments for asthma endotypes. Paediatr Respir Rev. 2020 Nov;36:118–27.

33. Niessen NM, Fricker M, McDonald VM, Gibson PG. T2-low: what do we know?: Past, present, and future of biologic therapies in noneosinophilic asthma. Ann Allergy Asthma Immunol. 2022 Aug;129(2):150–9.

34. Sporik R, Holgate ST, Platts-Mills TA, Cogswell JJ. Exposure to house-dust mite allergen (Der p I) and the development of asthma in childhood. A prospective study. N Engl J Med. 1990 Aug 23;323(8):502–7.

35. Permaul P, Hoffman EB, Fu C, Sheehan WJ, Baxi SN, Gaffin JM, et al. Allergens in Urban Schools and Homes of Children with Asthma. Pediatr Allergy Immunol. 2012 Sep;23(6):543–9.

36. Olivieri M, Zock JP, Accordini S, Heinrich J, Jarvis D, Künzli N, et al. Risk factors for new-onset cat sensitization among adults: a population-based international cohort study. J Allergy Clin Immunol. 2012 Feb;129(2):420–5.

37. Neill DR, Wong SH, Bellosi A, Flynn RJ, Daly M, Langford TKA, et al. Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature. 2010 Apr 29;464(7293):1367–70.

38. Kim HY, DeKruyff RH, Umetsu DT. The many paths to asthma: phenotype shaped by innate and adaptive immunity. Nat Immunol. 2010 Jul;11(7):577–84.

39. Holgate ST, Wenzel S, Postma DS, Weiss ST, Renz H, Sly PD. Asthma. Nat Rev Dis Primers. 2015;1(15025).

40. Djukanović R, Roche WR, Wilson JW, Beasley CR, Twentyman OP, Howarth RH, et al. Mucosal inflammation in asthma. Am Rev Respir Dis. 1990 Aug;142(2):434–57.

41. Truyen E, Coteur L, Dilissen E, Overbergh L, Dupont LJ, Ceuppens JL, et al. Evaluation of airway inflammation by quantitative Th1/Th2 cytokine mRNA measurement in sputum of asthma patients. Thorax. 2006 Mar;61(3):202–8.

42. Chesné J, Braza F, Mahay G, Brouard S, Aronica M, Magnan A. IL-17 in severe asthma. Where do we stand? Am J Respir Crit Care Med. 2014 Nov 15;190(10):1094–101.

43. Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet. 2011 Sep 10;378(9795):1006–14.

44. Gibson PG, Foster PS. Neutrophilic asthma: welcome back! Eur Respir J. 2019 Nov;54(5):1901846.

45. Östling J, van Geest M, Schofield JPR, Jevnikar Z, Wilson S, Ward J, et al. IL-17-high asthma with features of a psoriasis immunophenotype. J Allergy Clin Immunol. 2019 Nov;144(5):1198–213.

46. Hinks T, Zhou X, Staples K, Dimitrov B, Manta A, Petrossian T, et al. Multidimensional endotypes of asthma: topological data analysis of cross-

sectional clinical, pathological, and immunological data. Lancet. 2015 Feb 26;385 Suppl 1:S42.

47. Hinks TSC, Zhou X, Staples KJ, Dimitrov BD, Manta A, Petrossian T, et al. Innate and adaptive T cells in asthmatic patients: Relationship to severity and disease mechanisms. J Allergy Clin Immunol. 2015 Aug;136(2):323–33.

48. Mitchell AB, Glanville AR. The Human Respiratory Microbiome: Implications and Impact. Semin Respir Crit Care Med. 2018 Apr;39(2):199–212.

49. Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med. 2013 Jun;7(3):245–57.

50. Natalini JG, Singh S, Segal LN. The dynamic lung microbiome in health and disease. Nat Rev Microbiol. 2023 Apr;21(4):222–35.

51. Wu HJ, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. Gut Microbes. 2012 Jan 1;3(1):4–14.

52. Bosch AATM, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. EBioMedicine. 2016 Jul;9:336–45.

53. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial Topography of the Healthy Human Lower Respiratory Tract. mBio. 2017 Feb 14;8(1):e02287-16.

54. Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. Genome Med. 2013;5(7):63.

55. Caselli E, Fabbri C, D'Accolti M, Soffritti I, Bassi C, Mazzacane S, et al. Defining the oral microbiome by whole-genome sequencing and resistome analysis: the complexity of the healthy picture. BMC Microbiol. 2020 May 18;20(1):120.

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

57. Young JC, Chehoud C, Bittinger K, Bailey A, Diamond JM, Cantu E, et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. Am J Transplant. 2015 Jan;15(1):200–9.

58. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med. 2015 Apr 15;191(8):932–42.

59. Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol. 1977;31:107–33.

60. Hillman ET, Lu H, Yao T, Nakatsu CH. Microbial Ecology along the Gastrointestinal Tract. Microbes Environ. 2017 Dec 27;32(4):300–13.

61. Yang Y, Huang W, Fan Y, Chen GQ. Gastrointestinal Microenvironment and the Gut-Lung Axis in the Immune Responses of Severe COVID-19. Front Mol Biosci. 2021;8:647508.

62. Marsland BJ, Trompette A, Gollwitzer ES. The Gut-Lung Axis in Respiratory Disease. Ann Am Thorac Soc. 2015 Nov;12 Suppl 2:S150-156.

63. Chung KF. Airway microbial dysbiosis in asthmatic patients: A target for prevention and treatment? J Allergy Clin Immunol. 2017 Apr;139(4):1071–81.

64. Qin N, Zheng B, Yao J, Guo L, Zuo J, Wu L, et al. Influence of H7N9 virus infection and associated treatment on human gut microbiota. Sci Rep. 2015 Oct 22;5:14771.

65. Sokolowska M, Frei R, Lunjani N, Akdis CA, O'Mahony L. Microbiome and asthma. Asthma Res Pract. 2018 Jan 5;4:1.

66. Hu T, Dong Y, Yang C, Zhao M, He Q. Pathogenesis of Children's Allergic Diseases: Refocusing the Role of the Gut Microbiota. Front Physiol. 2021;12:749544.

67. Budden KF, Shukla SD, Rehman SF, Bowerman KL, Keely S, Hugenholtz P, et al. Functional effects of the microbiota in chronic respiratory disease. Lancet Respir Med. 2019 Oct;7(10):907–20.

68. Chung KF. Potential role of the lung microbiome in shaping asthma phenotypes. Ann Am Thorac Soc. 2017;14(5).

69. Park H, Shin JW, Park SG, Kim W. Microbial Communities in the Upper Respiratory Tract of Patients with Asthma and Chronic Obstructive Pulmonary Disease. Chu HW, editor. PLoS ONE. 2014 Oct 16;9(10):e109710.

70. Zhang Q, Cox M, Liang Z, Brinkmann F, Cardenas PA, Duff R, et al. Airway Microbiota in Severe Asthma and Relationship to Asthma Severity and Phenotypes. PLoS One. 2016;11(4):e0152724.

71. Wood LG, Simpson JL, Hansbro PM, Gibson PG. Potentially pathogenic bacteria cultured from the sputum of stable asthmatics are associated with increased 8-isoprostane and airway neutrophilia. Free Radical Research. 2010 Jan;44(2):146–54.

72. Hauptmann M, Schaible UE. Linking microbiota and respiratory disease. FEBS Lett. 2016 Nov;590(21):3721–38.

73. Martinez FD, Guerra S. Early Origins of Asthma. Role of Microbial Dysbiosis and Metabolic Dysfunction. Am J Respir Crit Care Med. 2018 Mar 1;197(5):573–9.

74. Stiemsma LT, Reynolds LA, Turvey SE, Finlay BB. The hygiene hypothesis: current perspectives and future therapies. Immunotargets Ther. 2015;4:143–57.

75. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bønnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med. 2007 Oct 11;357(15):1487–95.

76. Tang HHF, Lang A, Teo SM, Judd LM, Gangnon R, Evans MD, et al. Developmental patterns in the nasopharyngeal microbiome during infancy are associated with asthma risk. J Allergy Clin Immunol. 2021 May;147(5):1683–91.

77. Teo SM, Tang HHF, Mok D, Judd LM, Watts SC, Pham K, et al. Airway Microbiota Dynamics Uncover a Critical Window for Interplay of Pathogenic Bacteria and Allergy in Childhood Respiratory Disease. Cell Host Microbe. 2018 Sep 12;24(3):341-352.e5.

78. Patrick DM, Sbihi H, Dai DLY, Al Mamun A, Rasali D, Rose C, et al. Decreasing antibiotic use, the gut microbiota, and asthma incidence in children: evidence from population-based and prospective cohort studies. Lancet Respir Med. 2020 Nov;8(11):1094–105.

79. Murk W, Risnes KR, Bracken MB. Prenatal or early-life exposure to antibiotics and risk of childhood asthma: a systematic review. Pediatrics. 2011 Jun;127(6):1125–38.

80. Penders J, Kummeling I, Thijs C. Infant antibiotic use and wheeze and asthma risk: a systematic review and meta-analysis. Eur Respir J. 2011 Aug;38(2):295–302.

81. Ramirez J, Guarner F, Bustos Fernandez L, Maruy A, Sdepanian VL, Cohen H. Antibiotics as Major Disruptors of Gut Microbiota. Front Cell Infect Microbiol. 2020;10:572912.

82. Patangia DV, Anthony Ryan C, Dempsey E, Paul Ross R, Stanton C. Impact of antibiotics on the human microbiome and consequences for host health. Microbiologyopen. 2022 Feb;11(1):e1260.

83. Huang YJ, Nelson CE, Brodie EL, Desantis TZ, Baek MS, Liu J, et al. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. J Allergy Clin Immunol. 2011 Feb;127(2):372-381.e1-3.

84. Turek EM, Cox MJ, Hunter M, Hui J, James P, Willis-Owen SAG, et al. Airway microbial communities, smoking and asthma in a general population sample. EBioMedicine. 2021 Sep;71:103538.

85. Durack J, Lynch SV, Nariya S, Bhakta NR, Beigelman A, Castro M, et al. Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. J Allergy Clin Immunol. 2017 Jul;140(1):63–75.

86. Li N, Qiu R, Yang Z, Li J, Chung KF, Zhong N, et al. Sputum microbiota in severe asthma patients: Relationship to eosinophilic inflammation. Respir Med. 2017 Oct;131:192–8.

87. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, Neerincx AH, Riley JH, Bates S. Sputum microbiome profiles identify severe asthma phenotypes of relative stability at 12 to 18 months. Journal of Allergy and Clinical Immunology. 2021;147(1):123–34.

88. Durack J, Christian LS, Nariya S, Gonzalez J, Bhakta NR, Ansel KM, et al. Distinct associations of sputum and oral microbiota with atopic, immunologic, and clinical features in mild asthma. J Allergy Clin Immunol. 2020 Nov;146(5):1016–26.

89. Gibson PG, Yang IA, Upham JW, Reynolds PN, Hodge S, James AL, et al. Efficacy of azithromycin in severe asthma from the AMAZES randomised trial. ERJ Open Res. 2019 Oct;5(4):00056–2019.

90. Sverrild A, Kiilerich P, Brejnrod A, Pedersen R, Porsbjerg C, Bergqvist A, et al. Eosinophilic airway inflammation in asthmatic patients is associated with an altered airway microbiome. J Allergy Clin Immunol. 2017 Aug;140(2):407-417.e11.

91. Green BJ, Wiriyachaiporn S, Grainge C, Rogers GB, Kehagia V, Lau L, et al. Potentially pathogenic airway bacteria and neutrophilic inflammation in treatment resistant severe asthma. PLoS One. 2014;9(6):e100645.

92. Chotirmall SH, Bogaert D, Chalmers JD, Cox MJ, Hansbro PM, Huang YJ, et al. Therapeutic Targeting of the Respiratory Microbiome. Am J Respir Crit Care Med. 2022 Sep 1;206(5):535–44.

93. Global Strategy for Asthma Management and Prevention [Internet]. 2023 [cited 2023 Sep 12]. Available from: https://ginasthma.org/wp-content/uploads/2023/07/GINA-2023-Full-report-23_07_06-WMS.pdf

94. Papi A, Blasi F, Canonica GW, Morandi L, Richeldi L, Rossi A. Treatment strategies for asthma: reshaping the concept of asthma management. Allergy Asthma Clin Immunol. 2020;16:75.

95. Wilson SR, Strub P, Buist AS, Knowles SB, Lavori PW, Lapidus J, et al. Shared treatment decision making improves adherence and outcomes in poorly controlled asthma. Am J Respir Crit Care Med. 2010 Mar 15;181(6):566–77.

96. Barnes PJ. Inhaled Corticosteroids. Pharmaceuticals (Basel). 2010 Mar 8;3(3):514–40.

97. Chauhan BF, Ducharme FM. Anti-leukotriene agents compared to inhaled corticosteroids in the management of recurrent and/or chronic asthma in adults and children. Cochrane Database Syst Rev. 2012 May 16;2012(5):CD002314.

98. O'Byrne PM, FitzGerald JM, Bateman ED, Barnes PJ, Zhong N, Keen C, et al. Inhaled Combined Budesonide-Formoterol as Needed in Mild Asthma. N Engl J Med. 2018 May 17;378(20):1865–76.

99. Hardy J, Baggott C, Fingleton J, Reddel HK, Hancox RJ, Harwood M, et al. Budesonide-formoterol reliever therapy versus maintenance budesonide plus terbutaline reliever therapy in adults with mild to moderate asthma (PRACTICAL): a 52-week, open-label, multicentre, superiority, randomised controlled trial. Lancet. 2019 Sep 14;394(10202):919–28.

100. Beasley R, Holliday M, Reddel HK, Braithwaite I, Ebmeier S, Hancox RJ, et al. Controlled Trial of Budesonide-Formoterol as Needed for Mild Asthma. N Engl J Med. 2019 May 23;380(21):2020–30.

101. Agache I, Rocha C, Beltran J, Song Y, Posso M, Solà I, et al. Efficacy and safety of treatment with biologicals (benralizumab, dupilumab and omalizumab) for severe allergic asthma: A systematic review for the EAACI Guidelines - recommendations on the use of biologicals in severe asthma. Allergy. 2020 May;75(5):1043–57.

102. Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. N Engl J Med. 2009 Mar 5;360(10):973–84.

103. Castro M, Zangrilli J, Wechsler ME, Bateman ED, Brusselle GG, Bardin P, et al. Reslizumab for inadequately controlled asthma with elevated blood eosinophil counts: results from two multicentre, parallel, double-blind, randomised, placebo-controlled, phase 3 trials. Lancet Respir Med. 2015 May;3(5):355–66.

104. Bacharier LB, Maspero JF, Katelaris CH, Fiocchi AG, Gagnon R, de Mir I, et al. Dupilumab in Children with Uncontrolled Moderate-to-Severe Asthma. N Engl J Med. 2021 Dec 9;385(24):2230–40.

105. Castro M, Corren J, Pavord ID, Maspero J, Wenzel S, Rabe KF, et al. Dupilumab Efficacy and Safety in Moderate-to-Severe Uncontrolled Asthma. N Engl J Med. 2018 Jun 28;378(26):2486–96.

106. Wenzel S, Castro M, Corren J, Maspero J, Wang L, Zhang B, et al. Dupilumab efficacy and safety in adults with uncontrolled persistent asthma despite use of medium-to-high-dose inhaled corticosteroids plus a long-acting β2 agonist: a randomised double-blind placebo-controlled pivotal phase 2b dose-ranging trial. Lancet. 2016 Jul 2;388(10039):31–44.

107. Leitao Filho FS, Takiguchi H, Akata K, Ra SW, Moon JY, Kim HK, et al. Effects of Inhaled Corticosteroid/Long-Acting β2-Agonist Combination on the Airway

Microbiome of Patients with Chronic Obstructive Pulmonary Disease: A Randomized Controlled Clinical Trial (DISARM). Am J Respir Crit Care Med. 2021 Nov 15;204(10):1143–52.

108. Hartmann JE, Albrich WC, Dmitrijeva M, Kahlert CR. The Effects of Corticosteroids on the Respiratory Microbiome: A Systematic Review. Front Med (Lausanne). 2021;8:588584.

109. Loverdos K, Bellos G, Kokolatou L, Vasileiadis I, Giamarellos E, Pecchiari M, et al. Lung Microbiome in Asthma: Current Perspectives. J Clin Med. 2019 Nov 14;8(11):1967.

110. Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. Am J Respir Crit Care Med. 2013 Nov 15;188(10):1193–201.

111. Martin MJ, Zain NMM, Hearson G, Rivett DW, Koller G, Wooldridge DJ, et al. The airways microbiome of individuals with asthma treated with high and low doses of inhaled corticosteroids. PLoS One. 2020;15(12):e0244681.

112. Sullivan A, Ward C, Casey D, Flynn D, Hunt EB, Eustace JA, et al. Anti-IL-5 therapy and the airway microbiome in asthma. Eur Respir J. 2020 Sep 7;56(suppl 64).

113. Diver S, Haldar K, McDowell PJ, Busby J, Mistry V, Micieli C, et al. Relationship between inflammatory status and microbial composition in severe asthma and during exacerbation. Allergy. 2022 Nov;77(11):3362–76.

114. George L, Wright A, Mistry V, Sutcliffe A, Chachi L, Haldar K, et al. Sputum Streptococcus pneumoniae is reduced in COPD following treatment with benralizumab. Int J Chron Obstruct Pulmon Dis. 2019;14:1177–85.

115. Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, et al. Statistical Cluster Analysis of the British Thoracic Society Severe Refractory Asthma Registry: Clinical Outcomes and Phenotype Stability. PLoS One. 2014 Jul 24;9(7):e102987.

116. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. Genet Med. 2009 Jan;11(1):3–14.

117. Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. J Allergy Clin Immunol. 2011 Feb;127(2):355–60.

118. Aberg N, Hesselmar B, Aberg B, Eriksson B. Increase of asthma, allergic rhinitis and eczema in Swedish schoolchildren between 1979 and 1991. Clin Exp Allergy. 1995 Sep;25(9):815–9.

119. Sullivan PW, Ghushchyan VH, Slejko JF, Belozeroff V, Globe DR, Lin SL. The burden of adult asthma in the United States: evidence from the Medical Expenditure Panel Survey. J Allergy Clin Immunol. 2011 Feb;127(2):363-369.e1-3.

120. NHS England. Service specification number 170002/S. Specialised Respiratory Services (adult) – Severe Asthma [Internet]. 2017; [cited 2023 Dec 15]. Available from: https://www.england.nhs.uk/wp-content/uploads/2017/04/specialised-respiratory-services-adult-severe-asthma.pdf

121. Lewens T, McMillan J. Defining disease. Lancet. 2004 Feb 21;363(9409):664.

122. Muraro A, Lemanske RF, Hellings PW, Akdis CA, Bieber T, Casale TB, et al. Precision medicine in patients with allergic diseases: Airway diseases and atopic dermatitis-PRACTALL document of the European Academy of Allergy and Clinical Immunology and the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol. 2016 May;137(5):1347–58.

123. Boyd KM. Disease, illness, sickness, health, healing and wholeness: exploring some elusive concepts. Med Humanit. 2000 Jun;26(1):9–17.

124. Siddiqui S, Denlinger LC, Fowler SJ, Akuthota P, Shaw DE, Heaney LG, et al. Unmet Needs in Severe Asthma Subtyping and Precision Medicine Trials. Bridging Clinical and Patient Perspectives. Am J Respir Crit Care Med. 2019 Apr 1;199(7):823–9.

125. Leckie MJ, ten Brinke A, Khan J, Diamant Z, O'Connor BJ, Walls CM, et al. Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. Lancet. 2000 Dec 23;356(9248):2144–8.

126. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017 May 5;18(1):83.

127. O'Brien SJ. A decade of GigaScience: A perspective on conservation genetics. Gigascience. 2022 Jun 14;11:giac055.

128. Schriml LM, Lichenstein R, Bisordi K, Bearer C, Baron JA, Greene C. Modeling the enigma of complex disease etiology. J Transl Med. 2023 Feb 25;21(1):148.

129. Greene CS, Troyanskaya OG. Integrative systems biology for data-driven knowledge discovery. Semin Nephrol. 2010 Sep;30(5):443–54.

130. Li CX, Wheelock CE, Sköld CM, Wheelock ÅM. Integration of multi-omics datasets enables molecular classification of COPD. Eur Respir J. 2018 May;51(5):1701930.

131. Yan Z, Chen B, Yang Y, Yi X, Wei M, Ecklu-Mensah G, et al. Multi-omics analyses of airway host-microbe interactions in chronic obstructive pulmonary disease identify potential therapeutic interventions. Nat Microbiol. 2022 Sep;7(9):1361–75.

132. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.

133. Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA. Microbial ecology and evolution: a ribosomal RNA approach. Annu Rev Microbiol. 1986;40:337–65.

134. Simpson JL, Daly J, Baines KJ, Yang IA, Upham JW, Reynolds PN, et al. Airway dysbiosis: Haemophilus influenzae and Tropheryma in poorly controlled asthma. Eur Respir J. 2016 Mar;47(3):792–800.

135. Huang YJ, Nariya S, Harris JM, Lynch SV, Choy DF, Arron JR, et al. The airway microbiome in patients with severe asthma: Associations with disease features and severity. J Allergy Clin Immunol. 2015 Oct;136(4):874–84.

136. Marri PR, Stern DA, Wright AL, Billheimer D, Martinez FD. Asthma-associated differences in microbial composition of induced sputum. J Allergy Clin Immunol. 2013 Feb;131(2):346-352.e1-3.

137. Cait A, Hughes MR, Antignano F, Cait J, Dimitriu PA, Maas KR, et al. Microbiome-driven allergic lung inflammation is ameliorated by short-chain fatty acids. Mucosal Immunol. 2018 May;11(3):785–95.

138. Denner DR, Sangwan N, Becker JB, Hogarth DK, Oldham J, Castillo J, et al. Corticosteroid therapy and airflow obstruction influence the bronchial microbiome, which is distinct from that of bronchoalveolar lavage in asthmatic airways. J Allergy Clin Immunol. 2016 May;137(5):1398-1405.e3.

139. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019 Jun;20(6):341–55.

140. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel). 2023 Jul 13;12(7):997.

141. Noble D. A theory of biological relativity: no privileged level of causation. Interface Focus. 2012 Feb 6;2(1):55–64.

142. Athieniti E, Spyrou GM. A guide to multi-omics data collection and integration for translational medicine. Comput Struct Biotechnol J. 2023;21:134–49.

143. Kristensen VN, Lingjærde OC, Russnes HG, Vollan HKM, Frigessi A, Børresen-Dale AL. Principles and methods of integrative genomic analyses in cancer. Nat Rev Cancer. 2014 May;14(5):299–313.

144. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics. 2009 Nov 15;25(22):2906–12.

145. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. Proc Natl Acad Sci U S A. 2013 Mar 12;110(11):4245–50.

146. Meng C, Helm D, Frejno M, Kuster B. moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets. J Proteome Res. 2016 Mar 4;15(3):755–65.

147. Liu B, Shen X, Pan W. Integrative and regularized principal component analysis of multiple sources of data. Stat Med. 2016 Jun 15;35(13):2235–50.

148. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014 Mar;11(3):333–7.

149. Jiang X, Hu X. Inferring microbial interaction networks based on consensus similarity network fusion. Sci China Life Sci. 2014 Nov;57(11):1115–20.

150. National Research Council (US) Committee on A Framework for Developing a NewTaxonomy of Disease. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease [Internet]. Washington (DC): National Academies Press (US); 2011 [cited 2023 Aug 30]. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: http://www.ncbi.nlm.nih.gov/books/NBK91503/

151. Busse WW, Kraft M. Current unmet needs and potential solutions to uncontrolled asthma. Eur Respir Rev. 2022 Mar 31;31(163):210176.

152. Diamant Z, Vijverberg S, Alving K, Bakirtas A, Bjermer L, Custovic A, et al. Toward clinically applicable biomarkers for asthma: An EAACI position paper. Allergy. 2019 Oct;74(10):1835–51.

153. Chung KF. Personalised medicine in asthma: time for action: Number 1 in the Series “Personalised medicine in respiratory diseases” Edited by Renaud Louis and Nicolas Roche. European Respiratory Review. 2017 Sep 30;26(145):170064.

154. Chung KF. New treatments for severe treatment-resistant asthma: targeting the right patient. Lancet Respir Med. 2013 Oct;1(8):639–52.

155. Bunyavanich S, Schadt EE. Systems biology of asthma and allergic diseases: a multiscale approach. J Allergy Clin Immunol. 2015 Jan;135(1):31–42.

156. Pfeffer PE, Rupani H, De Simoni A. Bringing the treatable traits approach to primary care asthma management. Front Allergy. 2023;4:1240375.

157. McDonald VM, Fingleton J, Agusti A, Hiles SA, Clark VL, Holland AE, et al. Treatable traits: a new paradigm for 21st century management of chronic airway diseases: Treatable Traits Down Under International Workshop report. Eur Respir J. 2019 May;53(5):1802058.

158. Hiles SA, Gibson PG, Agusti A, McDonald VM. Treatable Traits That Predict Health Status and Treatment Response in Airway Disease. J Allergy Clin Immunol Pract. 2021 Mar;9(3):1255-1264.e2.

159. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. Am J Respir Crit Care Med. 2011;184(6):662–71.

160. Pienkowska K, Wiehlmann L, Tümmler B. Airway microbial metagenomics. Microbes Infect. 2018;20(9–10):536–42.

161. Taylor SL, Leong LEX, Choo JM. Inflammatory phenotypes in patients with severe asthma are associated with distinct airway microbiology. J Allergy Clin Immunol. 2018;141(1):94–103.

162. Martin MJ, Beasley R, Harrison TW. Towards a personalised treatment approach for asthma attacks. Thorax. 2020 Dec;75(12):1119–29.

163. Connolly C, Borg C, Gittins A, Shrimanker R, Hynes G, Hinks T, et al. P46 Treatable traits in a severe asthma cohort. Thorax. 2018 Dec 1;73(Suppl 4):A123–4.

164. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. Eur Respir J. 2015;46(5):1308–21.

165. Fleming L, Murray C, Bansal AT, Hashimoto S, Bisgaard H, Bush A, et al. The burden of severe asthma in childhood and adolescence: results from the paediatric U-BIOPRED cohorts. Eur Respir J. 2015 Nov;46(5):1322–33.

166. Pavord ID, Pizzichini MM, Pizzichini E, Hargreave FE. The use of induced sputum to investigate airway inflammation. Thorax. 1997 Jun;52(6):498–501.

167. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb 15;31(4):e15.

168. Bigler J, Boedigheimer M, Schofield JPR, Skipp PJ, Corfield J, Rowe A, et al. A Severe Asthma Disease Signature from Gene Expression Profiling of Peripheral Blood from U-BIOPRED Cohorts. Am J Respir Crit Care Med. 2017 May 15;195(10):1311–20.

169. Rossios C, Pavlidis S, Hoda U, Kuo CH, Wiegman C, Russell K, et al. Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. J Allergy Clin Immunol. 2018 Feb;141(2):560–70.

170. Versi A, Ivan FX, Abdel-Aziz MI, Bates S, Riley J, Baribaud F, et al. Haemophilus influenzae and Moraxella catarrhalis in sputum of severe asthma with inflammasome and neutrophil activation. Allergy. 2023 Jun 7;

171. Mendez R, Kesh K, Arora N, Di Martino L, McAllister F, Merchant N, et al. Microbial dysbiosis and polyamine metabolism as predictive markers for early detection of pancreatic cancer. Carcinogenesis. 2020 Jul 10;41(5):561–70.

172. Brandsma J, Goss VM, Yang X, Bakke PS, Caruso M, Chanez P, et al. Lipid phenotyping of lung epithelial lining fluid in healthy human volunteers. Metabolomics. 2018;14(10):123.

173. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959 Aug;37(8):911–7.

174. Kraemer S, Sanders G, Katilius E, King E, Bauer T, Zichi D, et al. SOMAscanTM v3.0: A Sensitive and Reproducible, Multiplex Proteomic Platform that Measures over 1100 Analytes in Complex Matrices. J Biomol Tech. 2014 May;25(Suppl):S8.

175. Schofield JPR, Burg D, Nicholas B, Strazzeri F, Brandsma J, Staykova D, et al. Stratification of asthma phenotypes by airway proteomic signatures. J Allergy Clin Immunol. 2019 Jul;144(1):70–82.

176. Keefe AD, Pai S, Ellington A. Aptamers as therapeutics. Nat Rev Drug Discov. 2010 Jul;9(7):537–50.

177. Kim CH, Tworoger SS, Stampfer MJ, Dillon ST, Gu X, Sawyer SJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. Sci Rep. 2018 May 30;8(1):8382.

178. Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med. 2010 Jul 19;8:68.

179. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

180. team T pandas development. pandas-dev/pandas: Pandas [Internet]. Zenodo; 2020. Available from: https://doi.org/10.5281/zenodo.3509134

181. Van Rossum G, Drake Jr FL. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam; 1995.

182. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene Set Enrichment Analysis Made Simple. Stat Methods Med Res. 2009 Dec;18(6):565–75.

183. Smyth GK. limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. Bioinformatics and Computational Biology Solutions Using R and Bioconductor [Internet]. New York, NY: Springer; 2005 [cited 2022 Sep 26]. p. 397–420. (Statistics for Biology and Health). Available from: https://doi.org/10.1007/0-387-29362-0_23

184. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

185. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

186. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

187. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50.

188. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013 Dec;14(1):1–15.

189. Zyla J, Marczyk M, Weiner J, Polanska J. Ranking metrics in gene set enrichment analysis: do they matter? BMC Bioinformatics. 2017 May 12;18(1):256.

190. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

191. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9.

192. Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT, et al. IL-13–induced airway mucus production is attenuated by MAPK13 inhibition. J Clin Invest. 2012 Dec 3;122(12):4555–68.

193. Björklund ÅK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol. 2016 Apr;17(4):451–60.

194. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun. 2005 Jun;6(4):319–31.

195. de Magalhães JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 2009 Apr 1;25(7):875–81.

196. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011 Jun 15;27(12):1739–40.

197. Liu YX, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell. 2021 May;12(5):315–30.

198. Kinoshita Y, Niwa H, Uchida-Fujii E, Nukada T. Establishment and assessment of an amplicon sequencing method targeting the 16S-ITS-23S rRNA operon for analysis of the equine gut microbiome. Sci Rep. 2021 Jun 4;11(1):11884.

199. Kamble A, Sawant S, Singh H. 16S ribosomal RNA gene-based metagenomics: A review. Biomedical Research Journal. 2020 Jan 1;7:5.

200. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017 Sep 12;35(9):833–44.

201. Gounot JS, Chia M, Bertrand D, Saw WY, Ravikrishnan A, Low A, et al. Genome-centric analysis of short and long read metagenomes reveals uncharacterized microbiome diversity in Southeast Asians. Nat Commun. 2022 Oct 13;13(1):6044.

202. Johansen J, Plichta DR, Nissen JN, Jespersen ML, Shah SA, Deng L, et al. Genome binning of viral entities from bulk metagenomics data. Nat Commun. 2022 Feb 18;13(1):965.

203. Santiago-Rodriguez TM, Hollister EB. Unraveling the viral dark matter through viral metagenomics. Front Immunol. 2022;13:1005107.

204. Nam NN, Do HDK, Loan Trinh KT, Lee NY. Metagenomics: An Effective Approach for Exploring Microbial Diversity and Functions. Foods. 2023 May 25;12(11):2140.

205. Roumpeka DD, Wallace RJ, Escalettes F, Fotheringham I, Watson M. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. Front Genet. 2017;8:23.

206. Wajid B, Anwar F, Wajid I, Nisar H, Meraj S, Zafar A, et al. Music of metagenomics-a review of its applications, analysis pipeline, and associated tools. Funct Integr Genomics. 2022 Feb;22(1):3–26.

207. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019 Aug 8;178(4):779–94.

208. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014 Mar 3;15(3):R46.

209. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biology. 2019 Nov 28;20(1):257.

210. Lu J, Rincon N, Wood DE, Breitwieser F, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc. 2022 Dec;17(12):2815–39.

211. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016 Apr 13;7:11257.

212. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012 Aug;9(8):811–4.

213. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018 Nov;15(11):962–8.

214. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D204–12.

215. Andrew S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online. 2010.

216. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.

217. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. The Sequence Alignment/Map format and SAMtools. Vol. 25. Bioinformatics (Oxford, England; 2009. p. 2078–9.

218. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

219. Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Vol. 25. Bioinformatics (Oxford, England; 2009. p. 2607–8.

220. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature methods. 2015;12(10):902–3.

221. Gweon HS, Shaw LP, Swann J, De Maio N, AbuOun M, Niehus R, et al. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ Microbiome. 2019 Oct 24;14(1):7.

222. Browne PD, Nielsen TK, Kot W, Aggerholm A, Gilbert MTP, Puetz L, et al. GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. Gigascience. 2020 Feb 1;9(2):giaa008.

223. Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, et al. Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. DNA Res. 2019 Oct 1;26(5):391–8.

224. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013 May 29;14(5):R51.

225. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre PM, Minchin D, et al. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. 2017.

226. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package [Internet]. 2022. Available from: https://CRAN.R-project.org/package=vegan

227. McMurdie PJ. Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013;8(4).

228. Liang Z, Zhang Q, Thomas CM, Chana KK, Gibeon D, Barnes PJ, et al. Impaired macrophage phagocytosis of bacteria in severe asthma. Respir Res. 2014 Jun 27;15(1):72.

229. Galdiero M, Galdiero M, Finamore E, Rossano F, Gambuzza M, Catania MR, et al. Haemophilus influenzae Porin Induces Toll-Like Receptor 2-Mediated Cytokine Production in Human Monocytes and Mouse Macrophages. Infect Immun. 2004 Feb;72(2):1204–9.

230. Clemans DL, Bauer RJ, Hanson JA, Hobbs MV, St. Geme JW, Marrs CF, et al. Induction of Proinflammatory Cytokines from Human Respiratory Epithelial Cells after Stimulation by Nontypeable Haemophilus influenzae. Infect Immun. 2000 Aug;68(8):4430–40.

231. Hassan F, Ren D, Zhang W, Merkel TJ, Gu XX. Moraxella catarrhalis activates murine macrophages through multiple toll like receptors and has reduced clearance in lungs from TLR4 mutant mice. PLoS One. 2012;7(5):e37610.

232. Fink J, Mathaba LT, Stewart GA, Graham PT, Steer JH, Joyce DA, et al. Moraxella catarrhalis stimulates the release of proinflammatory cytokines and prostaglandin E from human respiratory epithelial cells and monocyte-derived macrophages. FEMS Immunol Med Microbiol. 2006 Mar;46(2):198–208.

233. N'Guessan PD, Temmesfeld-Wollbrück B, Zahlten J, Eitel J, Zabel S, Schmeck B, et al. Moraxella catarrhalis induces ERK- and NF-kappaB-dependent COX-2 and prostaglandin E2 in lung epithelium. Eur Respir J. 2007 Sep;30(3):443–51.

234. Bousbia S, Papazian L, Auffray JP, Fenollar F, Martin C, Li W, et al. Tropheryma whipplei in Patients with Pneumonia. Emerg Infect Dis. 2010 Feb;16(2):258–63.

235. Lozupone C, Cota-Gomez A, Palmer BE, Linderman DJ, Charlson ES, Sodergren E, et al. Widespread colonization of the lung by Tropheryma whipplei in HIV infection. Am J Respir Crit Care Med. 2013 May 15;187(10):1110–7.

236. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013 May 15;187(10):1067–75.

237. Lim MY, Yoon HS, Rho M, Sung J, Song YM, Lee K, et al. Analysis of the association between host genetics, smoking, and sputum microbiota in healthy humans. Sci Rep. 2016 Mar 31;6:23745.

238. de Aquino SG, Abdollahi-Roodsaz S, Koenders MI, van de Loo FAJ, Pruijn GJM, Marijnissen RJ, et al. Periodontal pathogens directly promote autoimmune experimental arthritis by inducing a TLR2- and IL-1-driven Th17 response. J Immunol. 2014 May 1;192(9):4103–11.

239. Kim YH, Jang H, Kim SY, Jung JH, Kim GE, Park MR, et al. Gram-negative microbiota is related to acute exacerbation in children with asthma. Clin Transl Allergy. 2021 Oct;11(8):e12069.

240. Abdel-Aziz MI, Vijverberg SJH, Neerincx AH, Brinkman P, Wagener AH, Riley JH, et al. A multi-omics approach to delineate sputum microbiome-associated asthma inflammatory phenotypes. Eur Respir J. 2022 Jan;59(1):2102603.

241. Taylor SL, Leong LEX, Mobegi FM, Choo JM, Wesselingh S, Yang IA, et al. Long-Term Azithromycin Reduces Haemophilus influenzae and Increases Antibiotic Resistance in Severe Asthma. Am J Respir Crit Care Med. 2019 Aug 1;200(3):309–17.

242. Diver S, Richardson M, Haldar K, Ghebre MA, Ramsheh MY, Bafadhel M, et al. Sputum microbiomic clustering in asthma and chronic obstructive pulmonary disease reveals a Haemophilus-predominant subgroup. Allergy. 2020 Apr;75(4):808–17.

243. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003 Apr;4(2):249–64.

244. Morgat A, Coissac E, Coudert E, Axelsen KB, Keller G, Bairoch A, et al. UniPathway: a resource for the exploration and annotation of metabolic pathways. Nucleic Acids Res. 2012 Jan;40(Database issue):D761–9.

245. Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal. 1948;27(3):379–423.

246. Willis AD. Rarefaction, Alpha Diversity, and Statistics. Front Microbiol. 2019 Oct 23;10:2407.

247. Aitchison J. The Statistical Analysis of Compositional Data. (2).

248. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.

249. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987 Nov;20:53–65.

250. Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics. 1974 Jan 1;3(1):1–27.

251. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572–3.

252. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020 Jul 14;11(1):3514.

253. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I. A Transcriptome-driven Analysis of Epithelial Brushings and Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED. Am J Respir Crit Care Med. 2017;195(4):443–55.

254. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284–7.

255. Jevnikar Z, Östling J, Ax E, Calvén J, Thörn K, Israelsson E, et al. Epithelial IL-6 trans-signaling defines a new asthma phenotype with increased airway inflammation. J Allergy Clin Immunol. 2019 Feb;143(2):577–90.

256. Russell GJ, Brooks TM, McKinney MM, Anderson CG. Present and Future Taxonomic Selectivity in Bird and Mammal Extinctions. Conservation Biology. 1998;12(6):1365–76.

257. Purvis A, Agapow PM, Gittleman JL, Mace GM. Nonrandom extinction and the loss of evolutionary history. Science. 2000 Apr 14;288(5464):328–30.

258. Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, et al. Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. Projan SJ, editor. mBio. 2015 Dec 31;6(6):e01765-15.

259. Andrews CS, Miyata M, Susuki-Miyata S, Lee BC, Komatsu K, Li JD. Nontypeable Haemophilus influenzae-Induced MyD88 Short Expression Is Regulated by Positive IKKβ and CREB Pathways and Negative ERK1/2 Pathway. PLoS One. 2015;10(12):e0144840.

260. Uddin M, Watz H, Malmgren A, Pedersen F. NETopathic Inflammation in Chronic Obstructive Pulmonary Disease and Severe Asthma. Front Immunol. 2019 Feb 5;10:47.

261. Essilfie AT, Simpson JL, Horvat JC, Preston JA, Dunkley ML, Foster PS, et al. Haemophilus influenzae infection drives IL-17-mediated neutrophilic allergic airways disease. PLoS Pathog. 2011 Oct;7(10):e1002244.

262. Ackland J, Barber C, Heinson A, Azim A, Cleary DW, Christodoulides M, et al. Nontypeable Haemophilus influenzae infection of pulmonary macrophages drives neutrophilic inflammation in severe asthma. Allergy. 2022 Oct;77(10):2961–73.

263. Craig JE, Cliffe A, Garnett K, High NJ. Survival of nontypeable Haemophilus influenzae in macrophages. FEMS Microbiol Lett. 2001 Sep 11;203(1):55–61.

264. King P, Ngui J, Oppedisano F, Robins-Browne R, Holmes P, Holdsworth S. Effect of interferon gamma and CD40 ligation on intracellular monocyte survival of nontypeable Haemophilus influenzae. APMIS. 2008 Dec;116(12):1043–9.

265. de Vries SPW, Eleveld MJ, Hermans PWM, Bootsma HJ. Characterization of the molecular interplay between Moraxella catarrhalis and human respiratory tract epithelial cells. PLoS One. 2013;8(8):e72193.

266. Slevogt H, Schmeck B, Jonatat C, Zahlten J, Beermann W, Van Laak V, et al. *Moraxella catarrhalis* induces inflammatory response of bronchial epithelial cells via MAPK and NF-κB activation and histone deacetylase activity reduction. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2006 May;290(5):L818–26.

267. Rigauts C, Aizawa J, Taylor SL, Rogers GB, Govaerts M, Cos P, et al. R othia mucilaginosa is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. Eur Respir J. 2022 May;59(5):2101293.

268. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med. 2015 Sep 30;7(307):307ra152.

269. Brown MA, Jabeen M, Bharj G, Hinks TSC. Non-typeable Haemophilus influenzae airways infection: the next treatable trait in asthma? Eur Respir Rev. 2022 Sep 21;31(165):220008.

270. Hirano T, Matsunaga K. Late-onset asthma: current perspectives. J Asthma Allergy. 2018 Feb 9;11:19–27.

271. Zhu Z, Camargo CA, Raita Y, Freishtat RJ, Fujiogi M, Hahn A, et al. Nasopharyngeal airway dual-transcriptome of infants with severe bronchiolitis and risk of childhood asthma: A multicenter prospective study. J Allergy Clin Immunol. 2022 Oct;150(4):806–16.

272. Bisgaard H, Hermansen MN, Bønnelykke K, Stokholm J, Baty F, Skytt NL, et al. Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. BMJ. 2010 Oct 4;341:c4978.

273. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe. 2015 May 13;17(5):704–15.

274. Thorsen J, Rasmussen MA, Waage J, Mortensen M, Brejnrod A, Bønnelykke K, et al. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. Nat Commun. 2019 Nov 1;10(1):5001.

275. Toivonen L, Karppinen S, Schuez-Havupalo L, Waris M, He Q, Hoffman KL, et al. Longitudinal Changes in Early Nasal Microbiota and the Risk of Childhood Asthma. Pediatrics. 2020 Oct;146(4):e20200421.

276. Mansbach JM, Luna PN, Shaw CA, Hasegawa K, Petrosino JF, Piedra PA, et al. Increased Moraxella and Streptococcus species abundance after severe bronchiolitis is associated with recurrent wheezing. J Allergy Clin Immunol. 2020 Feb;145(2):518-527.e8.

277. Peroni DG, Nuzzi G, Trambusti I, Di Cicco ME, Comberiati P. Microbiome Composition and Its Impact on the Development of Allergic Diseases. Front Immunol. 2020;11:700.

278. Raita Y, Camargo CA, Bochkov YA, Celedón JC, Gern JE, Mansbach JM, et al. Integrated-omics endotyping of infants with rhinovirus bronchiolitis and risk of childhood asthma. J Allergy Clin Immunol. 2021 Jun;147(6):2108–17.

279. Raita Y, Pérez-Losada M, Freishtat RJ, Harmon B, Mansbach JM, Piedra PA, et al. Integrated omics endotyping of infants with respiratory syncytial virus bronchiolitis and risk of childhood asthma. Nat Commun. 2021 Jun 14;12(1):3601.

280. Rosas-Salazar C, Shilts MH, Tang ZZ, Hong Q, Turi KN, Snyder BM, et al. Exclusive breast-feeding, the early-life microbiome and immune response, and common childhood respiratory illnesses. J Allergy Clin Immunol. 2022 Sep;150(3):612–21.

281. Baan EJ, de Roos EW, Engelkes M, de Ridder M, Pedersen L, Berencsi K, et al. Characterization of Asthma by Age of Onset: A Multi-Database Cohort Study. J Allergy Clin Immunol Pract. 2022 Jul;10(7):1825-1834.e8.

282. Tan DJ, Walters EH, Perret JL, Burgess JA, Johns DP, Lowe AJ, et al. Clinical and functional differences between early-onset and late-onset adult asthma: a population-based Tasmanian Longitudinal Health Study. Thorax. 2016 Nov;71(11):981–7.

283. Faiz A, Heijink IH, Vermeulen CJ, Guryev V, van den Berge M, Nawijn MC, et al. Cigarette smoke exposure decreases CFLAR expression in the bronchial epithelium, augmenting susceptibility for lung epithelial cell death and DAMP release. Sci Rep. 2018 Aug 20;8(1):12426.

284. Brusselle G, Germinaro M, Weiss S, Zangrilli J. Reslizumab in patients with inadequately controlled late-onset asthma and elevated blood eosinophils. Pulm Pharmacol Ther. 2017 Apr;43:39–45.

285. Priya S, Burns MB, Ward T, Mars RAT, Adamowicz B, Lock EF, et al. Identification of shared and disease-specific host gene-microbiome associations

across human diseases using multi-omic integration. Nat Microbiol. 2022 Jun;7(6):780–95.

286. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014 Jan;15(1):34–48.

287. Narayana JK, Tsaneva-Atanasova K, Chotirmall SH. Microbiomics-focused Data Integration: A Fresh Solve for the Rubik's Cube of Endophenotyping? Am J Respir Crit Care Med. 2022 Aug 15;206(4):365–8.

288. Loxham M, Davies DE. Phenotypic and genetic aspects of epithelial barrier function in asthmatic patients. J Allergy Clin Immunol. 2017 Jun;139(6):1736–51.

289. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med. 2019 Jun;7(6):509–22.

290. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am J Hum Genet. 2019 Apr 4;104(4):665–84.

291. Azim A, Freeman A, Lavenu A, Mistry H, Haitchi HM, Newell C, et al. New Perspectives on Difficult Asthma; Sex and Age of Asthma-Onset Based Phenotypes. J Allergy Clin Immunol Pract. 2020;8(10):3396-3406.e4.

292. Heijink IH, Kuchibhotla VNS, Roffel MP, Maes T, Knight DA, Sayers I, et al. Epithelial cell dysfunction, a major driver of asthma development. Allergy. 2020 Aug;75(8):1902–17.

293. Sumitomo T, Nakata M, Nagase S, Takahara Y, Honda-Ogawa M, Mori Y, et al. GP96 Drives Exacerbation of Secondary Bacterial Pneumonia following Influenza A Virus Infection. mBio. 12(3):e03269-20.

294. Smith AK, Lange LA, Ampleford EJ, Meyers DA, Bleecker ER, Howard TD. Association of polymorphisms in CASP10 and CASP8 with FEV(1)/FVC and

bronchial hyperresponsiveness in ethnically diverse asthmatics. Clin Exp Allergy. 2008 Nov;38(11):1738–44.

295. Tan HTT, Hagner S, Ruchti F, Radzikowska U, Tan G, Altunbulakli C, et al. Tight junction, mucin, and inflammasome-related molecules are differentially expressed in eosinophilic, mixed, and neutrophilic experimental asthma in mice. Allergy. 2019 Feb;74(2):294–307.

296. Morales D, Micheva-Viteva S, Adikari S, Werner J, Wolinsky M, Hong-Geller E, et al. Targeting the Bet-Hedging Strategy with an Inhibitor of Bacterial Efflux Capacity Enhances Antibiotic Efficiency and Ameliorates Bacterial Persistence In Vitro. Microorganisms. 2022 Oct 5;10(10):1966.

297. Cosseau C, Devine DA, Dullaghan E, Gardy JL, Chikatamarla A, Gellatly S, et al. The commensal Streptococcus salivarius K12 downregulates the innate immune responses of human epithelial cells and promotes host-microbe homeostasis. Infect Immun. 2008 Sep;76(9):4163–75.

298. Pattaroni C, Watzenboeck ML, Schneidegger S, Kieser S, Wong NC, Bernasconi E, et al. Early-Life Formation of the Microbial and Immunological Environment of the Human Airways. Cell Host Microbe. 2018 Dec 12;24(6):857-865.e4.

299. Harris F, Berdugo YA, Tree T. IL-2-based approaches to Treg enhancement. Clin Exp Immunol. 2023 Mar 16;211(2):149–63.

300. Matchado MS, Rühlemann M, Reitmeier S, Kacprowski T, Frost F, Haller D, et al. On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. Microb Genom. 2024 Feb;10(2):001203.

301. Andreas S, Testa M, Boyer L, Brusselle G, Janssens W, Kerwin E, et al. Non-typeable Haemophilus influenzae-Moraxella catarrhalis vaccine for the prevention of exacerbations in chronic obstructive pulmonary disease: a multicentre, randomised, placebo-controlled, observer-blinded, proof-of-concept, phase 2b trial. Lancet Respir Med. 2022 May;10(5):435–46.

302. Winslow S, Odqvist L, Diver S, Riise R, Abdillahi S, Wingren C, et al. Multi-omics links IL-6 trans-signalling with neutrophil extracellular trap formation and *Haemophilus* infection in COPD. Eur Respir J. 2021 Mar 25;2003312.

303. Li R, Guo Q, Zhao J, Kang W, Lu R, Long Z, et al. Assessing causal relationships between gut microbiota and asthma: evidence from two sample Mendelian randomization analysis. Front Immunol. 2023;14:1148684.

304. Hinks TSC, Handley S, Keller B, Droit L, Staples KJ, Smith C, et al. S22 Analysis of the lung microbiome in human asthma using whole genome shot-gun metagenomics. Thorax. 2013 Dec 1;68(Suppl 3):A14–A14.

305. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. PLoS One. 2010 Dec 20;5(12):e15216.

306. Donaldson GC, Seemungal T a. R, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 2002 Oct;57(10):847–52.

307. Ying KL, Brasky TM, Freudenheim JL, McElroy JP, Nickerson QA, Song MA, et al. Saliva and Lung Microbiome Associations with Electronic Cigarette Use and Smoking. Cancer Prev Res (Phila). 2022 Jul 5;15(7):435–46.

308. Liu X, Sun W, Ma W, Wang H, Xu K, Zhao L, et al. Smoking related environmental microbes affecting the pulmonary microbiome in Chinese population. Sci Total Environ. 2022 Jul 10;829:154652.

309. Wang Z, Maschera B, Lea S, Kolsum U, Michalovich D, Van Horn S, et al. Airway host-microbiome interactions in chronic obstructive pulmonary disease. Respir Res. 2019 Jun 6;20(1):113.

310. Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nature Reviews Microbiology. 2017 Mar 20;15(5):259–70.

311. Goleva E, Hauk PJ, Hall CF, Liu AH, Riches DWH, Martin RJ, et al. Corticosteroid-resistant asthma is associated with classical antimicrobial activation of airway macrophages. J Allergy Clin Immunol. 2008 Sep;122(3):550-559.e3.

312. Chatkin JM, Dullius CR. The management of asthmatic smokers. Asthma Res Pract. 2016;2:10.

313. Lazarus SC, Chinchilli VM, Rollings NJ, Boushey HA, Cherniack R, Craig TJ, et al. Smoking affects response to inhaled corticosteroids or leukotriene receptor antagonists in asthma. Am J Respir Crit Care Med. 2007 Apr 15;175(8):783–90.

314. Thomson NC, Chaudhuri R, Heaney LG, Bucknall C, Niven RM, Brightling CE, et al. Clinical outcomes and inflammatory biomarkers in current smokers and exsmokers with severe asthma. J Allergy Clin Immunol. 2013 Apr;131(4):1008–16.

315. Chalmers GW, MacLeod KJ, Thomson L, Little SA, McSharry C, Thomson NC. Smoking and airway inflammation in patients with mild asthma. Chest. 2001 Dec;120(6):1917–22.

316. Telenga ED, Kerstjens HAM, Ten Hacken NHT, Postma DS, van den Berge M. Inflammation and corticosteroid responsiveness in ex-, current- and never-smoking asthmatics. BMC Pulm Med. 2013 Sep 22;13:58.

317. Broekema M, ten Hacken NHT, Volbeda F, Lodewijk ME, Hylkema MN, Postma DS, et al. Airway epithelial changes in smokers but not in ex-smokers with asthma. Am J Respir Crit Care Med. 2009 Dec 15;180(12):1170–8.

318. Boulet LP, Lemière C, Archambault F, Carrier G, Descary MC, Deschesnes F. Smoking and asthma: clinical and radiologic features, lung function, and airway inflammation. Chest. 2006 Mar;129(3):661–8.

319. Demarche S, Schleich F, Henket M, Paulus V, Van Hees T, Louis R. Detailed analysis of sputum and systemic inflammation in asthma phenotypes: are

paucigranulocytic asthmatics really non-inflammatory? BMC Pulmonary Medicine. 2016 Apr 5;16(1):46.

320. Jang AS, Park JS, Lee JH, Park SW, Kim DJ, Uh ST, et al. The impact of smoking on clinical and therapeutic effects in asthmatics. J Korean Med Sci. 2009 Apr;24(2):209–14.

321. Wu D, Lai T, Yuan Y, Chen M, Xia J, Li W, et al. Elevated expression of placental growth factor is associated with airway-wall vascular remodelling and thickening in smokers with asthma. Sci Rep. 2017 Feb 21;7:43017.

322. Klein DK, Silberbrandt A, Frøssing L, Hvidtfeldt M, von Bülow A, Nair P, et al. Impact of former smoking exposure on airway eosinophilic activation and autoimmunity in patients with severe asthma. Eur Respir J. 2022 Oct;60(4):2102446.

323. St-Laurent J, Bergeron C, Pagé N, Couture C, Laviolette M, Boulet LP. Influence of smoking on airway inflammation and remodelling in asthma. Clin Exp Allergy. 2008 Oct;38(10):1582–9.

324. Travers J, Rothenberg ME. Eosinophils in mucosal immune responses. Mucosal Immunol. 2015 May;8(3):464–75.

325. Bae JM. Interpretation of the hygiene and microflora hypothesis for allergic diseases through epigenetic epidemiology. Epidemiol Health. 2018 Mar 10;40:e2018006.

326. Lefaudeux D, De Meulder B, Loza MJ, Peffer N, Rowe A, Baribaud F, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. J Allergy Clin Immunol. 2017 Jun;139(6):1797–807.

327. Hekking PP, Loza MJ, Pavlidis S, de Meulder B, Lefaudeux D, Baribaud F, et al. Pathway discovery using transcriptomic profiles in adult-onset severe asthma. J Allergy Clin Immunol. 2018 Apr;141(4):1280–90.

328. Su YC, Singh B, Riesbeck K. Moraxella catarrhalis: from interactions with the host immune system to vaccine development. Future Microbiol. 2012 Sep;7(9):1073–100.

329. Prosperi MCF, Sahiner UM, Belgrave D, Sackesen C, Buchan IE, Simpson A, et al. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. Am J Respir Crit Care Med. 2013 Dec 1;188(11):1303–12.

330. Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. Metabolites. 2021 Dec 24;12(1):14.

331. Fang X, Lloyd CJ, Palsson BO. Reconstructing organisms in silico: genome-scale models and their emerging applications. Nat Rev Microbiol. 2020 Dec;18(12):731–43.

332. Lam TJ, Stamboulian M, Han W, Ye Y. Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. PLoS Comput Biol. 2020 Oct;16(10):e1007951.

333. Bonamichi-Santos R, Aun MV, Agondi RC, Kalil J, Giavina-Bianchi P. Microbiome and Asthma: What Have Experimental Models Already Taught Us? J Immunol Res. 2015;2015:614758.

334. Zosky GR, Larcombe AN, White OJ, Burchell JT, Janosi TZ, Hantos Z, et al. Ovalbumin-sensitized mice are good models for airway hyperresponsiveness but not acute physiological responses to allergen inhalation. Clin Exp Allergy. 2008 May;38(5):829–38.

335. Kips JC, Anderson GP, Fredberg JJ, Herz U, Inman MD, Jordana M, et al. Murine models of asthma. Eur Respir J. 2003 Aug;22(2):374–82.

336. Chang D, Sharma L, Dela Cruz CS. Harnessing Murine Microbiome Models to Study Human Lung Microbiome. Chest. 2020 Apr;157(4):776–8.

Chapter 9. Appendix

Appendix I. Gene signatures used in this thesis.

#	Name	Code Name	Long Name	Description	Genes	Reference (URL)
1	Lung-tissue resident Mφ	MACROPHAGES_TR	Tissue-resident Macrophage	tissue-resident alveolar macrophages gene expression vs monocyte derived alveolar macrophage differentially expressed gene	*CEBPB TGFBI CD36 TNF CD80 TOMM20 RNASEH2A PPARG RXRA APOA1BP FAM100A CSF2RA NDUFA4 RAC1 CAD SIGLEC1 SIRPA*	https:// pubmed.ncbi.nlm.nih.gov/ 28694385/
2	Neutrophil	NEUTROPHIL		Heparanized blood mixed with an equal volume of dextran/saline solution.	*ABTB1 AMPD2 FAM53C CCR3 CDA CMTM2 CLC CREB5 CTBS CTBS DICER1 MSL1 FCGR2B FCGR3B FCGR3B MANSC1 FPR2 FRAT2 GPR27 FFAR2 HSPA6 CXCR1 CXCR2 TMCC1 SLC45A4 KRT23 MBOAT7 TSEN34 MXD1 YPEL3 CEP19 R3HDM4 MPPE1 SLC25A37 NCF4 NRBF2 PHC2 PROK2 RALB RNF141 SEC14L1 MSRB1 STX3 TSPAN16 VMP1 VNN2 XPO6*	http:// www.ncbi.nlm.nih.gov/ pubmed/15789058
3	ILC2	ILC2	Small intestine lamina propria (*Mus musculus*)	IL2 genes uniquely upregulated at least 2-fold compared to ILC1, ILC3, and LTi-like	*PTGDR2 IL1RL1 IL17RB KLRG1 GATA3 IL13 RORA HLF PPARG MAF KLF7 ZEB2 POU1F1 HPGDS HPGD PTGER2 TCF7 NOTCH1 TLE4 FCRL3 IL10RA IL31*	http:// www.ncbi.nlm.nih.gov/ pubmed/25621825

					SLAMF1 P2RY1 TNFRSF9 ITGAM FASLG TNFRSF19	
4	IL13 T2	IL13_T2		IL-13–driven mucus production in human airway epithelial cells.	*CST1 CCL26 PRB2 PRB1 PRB3 POSTN PRB4 ITLN1 ALOX15 SH2D1B CA2 NOS2 FCGBP FOXA3 SPDEF CAPN14 DUOXA2 CLDN5 PADI3 TSPAN8 ALPL KCNJ16 FETUB B3GNT6 CDH26 LRRC31 MUC13 VSIG2 CSTA FAM3B SLC9B2 NTRK1 KLF4 HPDL SOCS1 TRNP1 HS3ST1 VWF DUOX2 CISH ATP13A5 ZNF808 RNASE4 CCBL1 SDCBP2 TMPRSS2 HYAL1 CCDC109B FAM83D TRAK1 TPK1 SLC7A1 CYP2C18 CDC42EP5 KCNS3 ADRA2A MRAP2 SLC2A10 PPARG FAM26E ADCY4 WNT3 SLCO4A1 ALDH1A2 C10orf99 WDFY2*	https:// pubmed.ncbi.nlm.nih.gov/ 23187130/

5	Oxidative Phosphorylation	OXPHOS		Genes encoding proteins involved in oxidative phosphorylation in MSigDB	*OXPHOS, ND1, ND2, ND3, ND4, ND4L, ND5, ND6, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS5, NDUFS6, NDUFS7, NDUFS8, NDUFV1, NDUFV2, NDUFV3, NDUFA1, NDUFA2, NDUFA3, NDUFA4, NDUFA4L2, NDUFA5, NDUFA6, NDUFA7, NDUFA8, NDUFA9, NDUFA10, NDUFAB1, NDUFA11, NDUFA12, NDUFA13, NDUFB1, NDUFB2, NDUFB3, NDUFB4, NDUFB5, NDUFB6, NDUFB7, NDUFB8, NDUFB9, NDUFB10, NDUFB11, NDUFC1, NDUFC2, NDUFC2-KCTD14, SDHA, SDHB, SDHC, SDHD, UQCRFS1, CYTB, CYC1, UQCRC1, UQCRC2, UQCRH, UQCRHL, UQCRB, UQCRQ, UQCR10, UQCR11, COX10, COX3, COX1, COX2, COX4I2, COX4I1, COX5A, COX5B, COX6A1, COX6A2, COX6B1, COX6B2, COX6C, COX7A1, COX7A2, COX7A2L, COX7B, COX7B2, COX7C, COX8C, COX8A, COX11, COX15, COX17, ATP5A1, ATP5B, ATP5C1, ATP5D, ATP5E, ATP5O, ATP6, ATP5F1, ATP5G1, ATP5G2, ATP5G3, ATP5H, ATP5I, ATP5J2, ATP5L, ATP5J, ATP8, ATP6V1A, ATP6V1B1, ATP6V1B2, ATP6V1C2, ATP6V1C1, ATP6V1D, ATP6V1E2, ATP6V1E1,*	https:// pubmed.ncbi.nlm.nih.gov/ 21546393/

					ATP6V1F, ATP6V1G1, ATP6V1G3, ATP6V1G2, ATP6V1H, TCIRG1, ATP6V0A2, ATP6V0A4, ATP6V0A1, ATP6V0C, ATP6V0B, ATP6V0D1, ATP6V0D2, ATP6V0E1, ATP6V0E2, ATP6AP1, ATP4A, ATP4B, ATP12A, PPA2, PPA1, LHPP	
6	NETosis	NETOSIS			*ITGAM C5AR1 MPO ELANE CTSG AGER CEACAM8 HMGB1 CCL*	https://pubmed.ncbi.nlm.nih.gov/34775962/
7	IL6 Transsignaling	IL6_TRANSSIGNALLING		bronchial epithelial cells) ALI (air liquid interface) cultures stimulated with IL6 to induce a gene profile.	*TNFAIP6 PDE4B IL1R2 S100A9 S100A8 S100A12 CHI3L1 SPP1*	https://pubmed.ncbi.nlm.nih.gov/29902480/
8	Blood Eosinophil	BLOOD_EOS_UP_PLOS2015			*DEFB105B BCKDHA ISG20 KRT23 VSIG10 CREB3L3 IL2RA IL10RA IL3RA ENST00000367534 PMP22 GSTT1 TMEM176B CCS S100A10 ACOT4 LIPA CCDC146 ENST00000449914 HRK CHRM4*	https://pubmed.ncbi.nlm.nih.gov/26524763/

9	Th17	TH17			*KLRB1 RORC PLXND1 CTSH ALOX5 PTPN13 IL4I1 C11orf75 NEFL HLF JAKMIP2 DSE LIMS1 HLA-DRB1 LTK HLA-DRB4 USP10 NR1D1 LCAT SAMD3 HSPG2*	https://pubmed.ncbi.nlm.nih.gov/23870669/
10	Aging	AGING		Meta-analysis of aging gene expression datasets (from mice, rats and human) to find common genes associated with aging across datasets	*MMACHC, PDE7B, CTSS, HLA-DRA, LUZP1, C3, C1QB, BRINP3, C1orf210, DENND6B, APOD, KHDRBS2, DHDDS, VWF, GPER1, CALHM2, MPEG1, FCGR2A, GPNMB, CLASP2, MSL3, C4A, MGST1, SHARPIN, APPBP2, AIP, IGJ, RNASET2, FCGR2B, ANTXR1, HIST1H1C, C1QA, RAB40B, CD74, LYZ, HMGN2, TLX3, SRPR, RORB, GFAP, ARRB1, MT2A, PTBP2, ABCB6, ARL11, KITLG, MMP10, UGT2B17, FXYD1, ANXA3, BIRC7, CDKN1A, AMH, NPC2, SH3GLB1, HBB, PCSK6, GSTA1*	https://pubmed.ncbi.nlm.nih.gov/19189975/
11	Inflammasome	INFLAMMASOME		Gene expression of Sputum of neutrophilic asthmatics vs healthy	*IL1B NLRP3 CASP1 CASP4 CASP5*	https://pubmed.ncbi.nlm.nih.gov/24136334/

Appendix II. Permissions for reproduction of materials

Figure 1

Licensing

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 2 (Full agreement available on request)

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

May 12, 2024

This Agreement between Mr. Ali Versi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5782641181533
License date	May 05, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Clinical Reviews in Allergy and Immunology
Licensed Content Title	Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease
Licensed Content Author	Merin E. Kuruvilla et al
Licensed Content Date	Sep 11, 2018
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations

Figure 3

Dear Ali,

Thank you for your permission enquiry.

European Respiratory Society hereby grants you permission to reproduce the material as requested for your dissertation/thesis (print and digital) and full acknowledgement must be given.

Citation:	Asthma management in low and middle income countries: case for change Kevin Mortimer, Helen K. Reddel, Paulo M. Pitrez, Eric D. Bateman European Respiratory Journal Sep 2022, 60 (3) 2103179; **DOI:** 10.1183/13993003.03179-2021
Material:	Figure 1 and Table 3
Acknowledgement Wording:	Reproduced with permission of the ERS 2024: European Respiratory Journal Sep 2022, 60 (3) 2103179; **DOI:** 10.1183/13993003.03179-2021

Copyright remains with ERS 2024. Unless otherwise stated the authors have granted ERS exclusive rights to licence this content. These publications are copyrighted material and must not be copied, reproduced, transferred, distributed, leased, licensed, placed in a storage retrieval system or publicly performed or used in any way except as specifically permitted in writing by the publishers (ERS), as allowed under the terms and conditions of which it was purchased or as strictly permitted by applicable copyright law. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given. Altering/Modifying Material: This is not permitted; however, figures and illustrations may be altered/adapted minimally to serve your work. This agreement is personal to you and may not be sublicense, assigned or transferred by you to any other person without our written permission. Any unauthorised distribution or use of this text may be a direct infringement of the publisher's rights and those responsible may be liable in law accordingly. This permission is for non-exclusive world English rights only. The ERS on this occasion does not allow permission for non-exclusive world rights in all other languages and future revisions and editions.

Figure 4

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 12, 2024

This Agreement between Mr. Ali Versi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5782650531965
License date	May 05, 2024
Licensed Content Publisher	Elsevier
Licensed Content Publication	Drug Discovery Today: Disease Models
Licensed Content Title	Modelling airway remodelling in asthma
Licensed Content Author	Clare M. Lloyd,Douglas S. Robinson,Sarah J. McMillan
Licensed Content Date	Dec 1, 2004
Licensed Content Volume	1
Licensed Content Issue	4
Licensed Content Pages	6
Start Page	425
End Page	430
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations

Figure 5

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 12, 2024

This Agreement between Mr. Ali Versi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5782661456856
License date	May 05, 2024
Licensed Content Publisher	Elsevier
Licensed Content Publication	Annals of Allergy, Asthma & Immunology
Licensed Content Title	T2-low: what do we know? Past, present, and future of biologic therapies in noneosinophilic asthma
Licensed Content Author	Natalie M. Niessen,Michael Fricker,Vanessa M. McDonald,Peter G. Gibson
Licensed Content Date	Aug 1, 2022
Licensed Content Volume	129
Licensed Content Issue	2
Licensed Content Pages	10
Start Page	150
End Page	159
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations

Figure 6

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

May 12, 2024

This Agreement between Mr. Ali Versi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5776891462391
License date	Apr 27, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Disease Primers
Licensed Content Title	Asthma
Licensed Content Author	Stephen T. Holgate et al
Licensed Content Date	Sep 10, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations

Figure 7

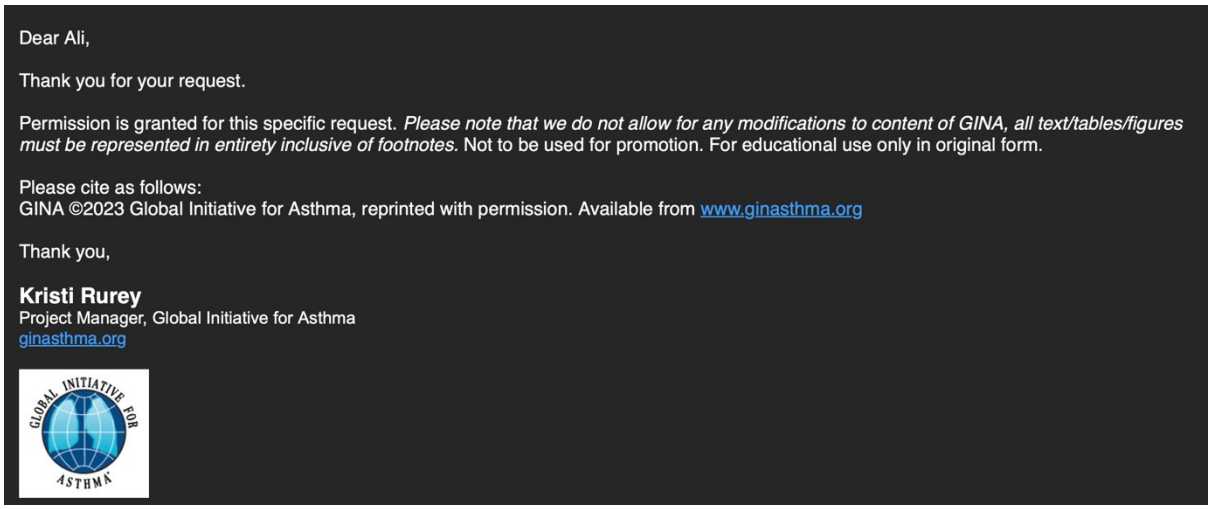
ELSEVIER LICENSE
TERMS AND CONDITIONS

May 12, 2024

This Agreement between Mr. Ali Versi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5771250002363
License date	Apr 17, 2024
Licensed Content Publisher	Elsevier
Licensed Content Publication	The Lancet Respiratory Medicine
Licensed Content Title	Functional effects of the microbiota in chronic respiratory disease
Licensed Content Author	Kurtis F Budden,Shakti D Shukla,Saima Firdous Rehman,Kate L Bowerman,Simon Keely,Philip Hugenholtz,Darius P H Armstrong-James,Ian M Adcock,Sanjay H Chotirmall,Kian Fan Chung,Philip M Hansbro
Licensed Content Date	Oct 1, 2019
Licensed Content Volume	7
Licensed Content Issue	10
Licensed Content Pages	14

Figure 8



Dear Ali,

Thank you for your request.

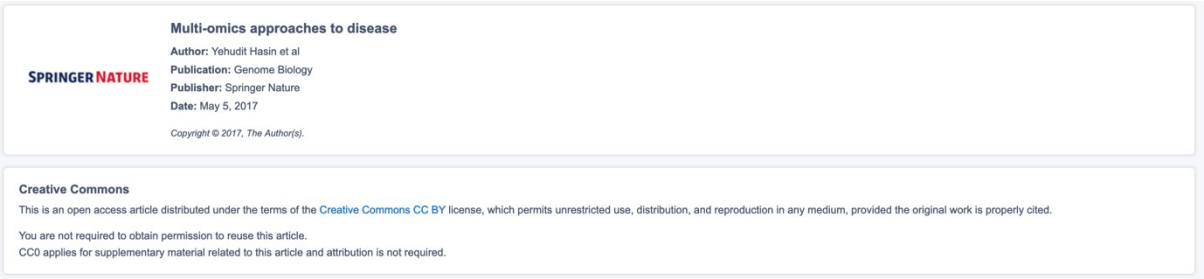
Permission is granted for this specific request. *Please note that we do not allow for any modifications to content of GINA, all text/tables/figures must be represented in entirety inclusive of footnotes.* Not to be used for promotion. For educational use only in original form.

Please cite as follows:
GINA ©2023 Global Initiative for Asthma, reprinted with permission. Available from www.ginasthma.org

Thank you,

Kristi Rurey
Project Manager, Global Initiative for Asthma
ginasthma.org

Figure 9



SPRINGER NATURE

Multi-omics approaches to disease

Author: Yehudit Hasin et al
Publication: Genome Biology
Publisher: Springer Nature
Date: May 5, 2017

Copyright © 2017, The Author(s).

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.
CC0 applies for supplementary material related to this article and attribution is not required.