

101034339 – PROMISE

Preparing for RSV Immunisation and Surveillance in Europe

WP4 – Task 4.2 Biomarker Identification Validation in a Prospective  
Observational Case-Control Study

## D4.7 Scientific report of combined multi-omics analysis

<b>Lead</b>	Jim Janimak (18 – GSK) <a href="mailto:Jim.j.janimak@gsk.com">Jim.j.janimak@gsk.com</a>
<b>Contributors</b>	Rodrigo Bacigalupe (18 – GSK) Martina Poletti (18 – GSK) Sarah Pötgens (18 – GSK)
<b>Reviewers</b>	Louis Bont (2 – UMCU); Ryan Thwaites (4 – Imperial); Rolf Kramer (17 – SP); Steven Sijmons (18 – GSK)

### Document History

Version	Date	Description
V0.1	19/04/2024	First Draft
V0.2	08/05/2024	Comments
V0.3	13/05/2024	Draft
V1.0	23/05/2024	Final Version

Reproduction of this document or part of this document without PROMISE Consortium permission is forbidden. Any use of any part must acknowledge the PROMISE Consortium as “This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under Grant Agreement 101034339. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA”. This document is shared within the PROMISE Consortium and is in line with the general communication guidelines described in the PROMISE Consortium Agreement.

## Table of contents

Table of contents.....	2
Definitions .....	3
Abbreviations .....	4
Abstract.....	5
1. Introduction.....	6
1.1. Background .....	6
1.2. Context of the analysis.....	6
1.3. Objectives and key questions .....	7
2. Methods.....	9
2.1. Data generation .....	9
2.2. Bioinformatics data preprocessing .....	10
2.3. Statistical analysis .....	11
2.4. Data retention and availability .....	13
3. Results.....	14
3.1. Description of metadata variables.....	14
3.2. Respiratory microbiome .....	15
3.3. Stool microbiome .....	20
3.4. Serum metabolome .....	23
3.5. Microbiome-metabolome associations .....	28
4. Discussion .....	30
5. Conclusion and next steps .....	33
6. Repository for primary data .....	33
7. References .....	34
ANNEXES.....	37
ANNEX I. Respiratory microbiome in RSV.....	37
ANNEX II. Microbiome-metabolite correlations in RSV .....	40
ANNEX III. Serum metabolomics analyses.....	42

## Definitions

- **Participants** of the PROMISE Consortium are referred to herein according to the following codes:
  1. **UEDIN.** The University of Edinburgh (United Kingdom)
  2. **UMCU.** Universitair Medisch Centrum Utrecht (Netherlands)
  3. **UA.** Universiteit Antwerpen (Belgium)
  4. **Imperial.** Imperial College of Science, Technology and Medicine (United Kingdom)
  5. **UOXF.** The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom)
  6. **THL.** Terveysten Ja Hyvinvoinnin Laitos (Finland)
  7. **RIVM.** Rijksinstituut voor Volksgezondheid en Milieu (Netherlands)
  8. **NIVEL.** Stichting Nedelands Instituut voor Onderzoek van de Gezondheidszorg (Netherlands)
  9. **TUCH.** Varsinais-Suomen Sairaanhoidopiirin Kuntayhtymä (Finland)
  10. **TEAMIT.** TEAM IT Research, S.L. (Spain)
  11. **ReSViNET.** Stichting Resvinet (Netherlands)
  12. **SSI.** Statens Serum Institut (Denmark)
  13. **SERGAS.** Servizo Galego de Saúde (Spain)
  14. **PENTA.** Fondazione PENTA - For the treatment and care of children with HIV and related diseases - ONLUS (Italy)
  15. **FISABIO.** Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (Spain)
  16. **MLU.** Martin-Luther-Universitaet Halle-Wittenberg (Germany)
  17. **SP.** Sanofi Pasteur, S.A. (France)
  18. **GSK.** GlaxoSmithKline Biologicals, S.A. (Belgium)
  19. **JANSSEN.** Janssen Pharmaceutica, N.V (Belgium)
  20. **Novavax.** Novavax, Inc. (United States)
  21. **Pfizer.** Pfizer Limited (United Kingdom)
  22. **AZ.** AstraZeneca AB (Sweden)
  
- **Grant Agreement.** (Including its annexes and any amendments) The agreement signed between the beneficiaries of the action and the IMI2 JU for the undertaking of the PROMISE project (Grant Agreement No. 101034339).
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium.** The PROMISE Consortium, comprising the above-mentioned participants.
- **Consortium Agreement.** The agreement concluded amongst PROMISE participants for the implementation of the Grant Agreement. The agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

## Abbreviations

Acronym / Abbreviation	Meaning
<b>16S rRNA</b>	16S ribosomal RNA
<b>CLR</b>	Centred log-ratio
<b>dbRDA</b>	Distance-based redundancy analysis
<b>FDR</b>	False discovery rate
<b>GTDB</b>	Genome Taxonomy Database
<b>GSK</b>	GlaxoSmithKline Biologicals
<b>HUMAnN</b>	HMP Unified Metabolic Analysis Network
<b>IQR</b>	Interquartile range
<b>LC-MS/MS</b>	Liquid chromatography-Mass spectrometry/Mass spectrometry
<b>Maaslin2</b>	Microbiome Multivariable Association with Linear Models 2
<b>PCA</b>	Principal component analysis
<b>PCoA</b>	Principal Coordinate Analysis
<b>PLS-DA</b>	Partial least square discriminant analysis
<b>PERMANOVA</b>	Permutational analysis of variance
<b>PROMISE</b>	Preparing for RSV Immunisation and Surveillance in Europe
<b>QC</b>	Quality control
<b>RSV</b>	Respiratory syncytial virus
<b>RESCEU</b>	REspiratory Syncytial virus Consortium in EUrope
<b>UPLC-MS/MS</b>	Ultra-high performance liquid chromatography-tandem mass spectroscopy
<b>UMCU</b>	Universitair Medisch Centrum Utrecht
<b>WGS</b>	Whole genome sequencing

## Abstract

In this study, we identified respiratory nasopharyngeal and stool microbiome and serum metabolome signatures of respiratory syncytial virus (RSV) infection severity. Previously healthy infants with different severities of RSV infection (medically attended non-hospitalised, hospitalised non-ventilated and hospitalised ventilated infants) were enrolled and compared to healthy controls. Microbiomics and metabolomics analyses were performed, including shotgun metagenomics of stool samples, 16S rRNA gene amplicon sequencing of nasopharyngeal samples and metabolomics profiling of serum samples. While stool shotgun metagenomics and serum metabolomics were performed only for samples taken at enrolment, respiratory tract 16S rRNA gene amplicon sequencing was obtained from samples taken at the first visit and at 7 weeks ( $\pm 1$  week) after discharge (convalescent time point; RSV-positive infants only). We identified that RSV infection significantly impacts the respiratory microbiome and serum metabolome, with minimal effect on the gut microbiome. We observed an increase in richness and shifts in the composition of the respiratory microbiome during RSV infection, with contrasting changes noted in recovered RSV infants. The study identified several co-variates (age, season, and breastfeeding) that influenced variations in the microbiome as well as the serum metabolites profiles. Despite these findings, the study faced limitations such as the lack of age-matched samples and seasonality differences between groups. In conclusion, while our study could not identify any marker of disease susceptibility or risk of RSV infection, it provides valuable insights into the impact of RSV on the microbiome and metabolome of infants. These findings could guide future research and therapeutic approaches for RSV.

# 1. Introduction

## 1.1. Background

Human respiratory syncytial virus (RSV) causes severe disease in young infants, with an estimated 26,000 in-hospital deaths worldwide in children <5 years old [1]. Besides young age, prematurity and congenital cardiopulmonary disease are known risk factors for severe disease. However, most children admitted to the hospital with RSV were previously healthy, and therefore biomarkers for disease severity are a key requirement to identify infants at risk for clinical management [2]. The Respiratory Syncytial virus Consortium in Europe (RESCEU) conducted a biomarker discovery study in healthy young infants who had RSV infection in 2 prospective studies (RESCEU infant cohort and infant case-control studies [3], [4] and identified biomarkers predictive of, or associated with, lower respiratory tract RSV infection and disease severity. Those biomarkers were identified by (1) a systematic literature review [5] and (2) an analysis of stool and respiratory microbiomes (Öner *et al.*, *in preparation*).

A critical step in biomarker development is external validation [6], required for clinical implementation and for regulatory requirements. In PROMISE (Preparing for RSV Immunisation and Surveillance in Europe), the overall goal of Work Package (WP) 4 was to validate biomarkers previously identified in RESCEU. This WP was divided in three major activities, with Task 4.2 aiming to validate the identified gut and respiratory microbiome biomarkers associated with RSV infection severity and respiratory sequelae. Microbiomes were studied in RESCEU because human-associated microbial components (specific bacterial taxa and metabolites) can locally and systemically influence the magnitude of the induced inflammatory responses and disease progression. As such, among the upcoming RESCEU results, a descriptive analysis of the gut and lung microbiomes in the infant cohorts generated an initial dataset to explore the interactions between the microbiome and RSV infection severity.

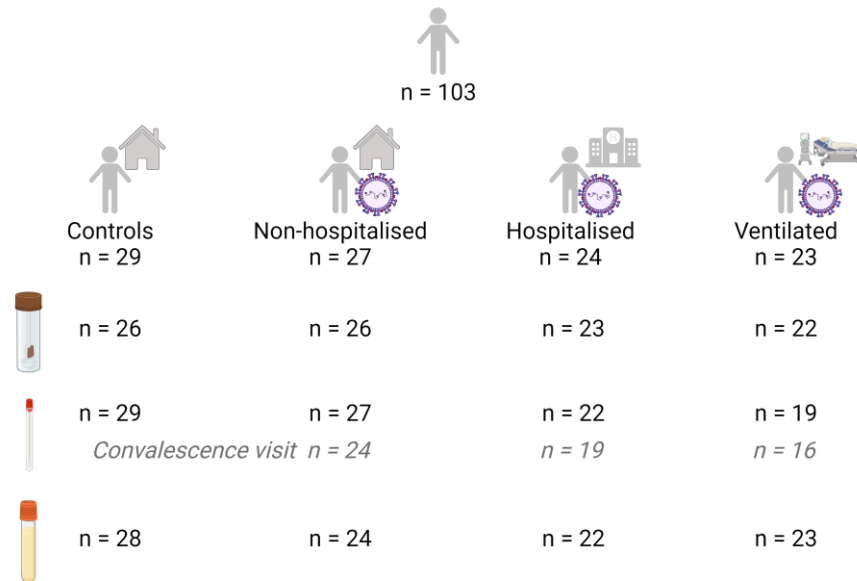
In the context of RSV, the observed inflammatory state at the respiratory mucosa could be mediated by signals originating from the gut microbiota [7]. To address this point, the PROMISE infant case-control cohort study collected paired gut and respiratory samples to perform metagenomics in addition to serum samples for metabolomics analyses (Report 4.1). This dual-omics approach allows the exploration of the associations between microbial components at the systemic and mucosal level and RSV infection severity. These results can contribute to our understanding of the mechanisms underlying the variable outcome of RSV exposure and could pave the way for the development of diagnostic tests to better predict RSV susceptibility in paediatric populations.

## 1.2. Context of the analysis

The guiding principle of this study is to build on RESCEU's WP5 achievements by moving beyond single descriptive studies and explore additional relationships between previously identified biomarkers (in RESCEU and from recent publications) and RSV infection disease outcomes. Samples from the respiratory tract, the stool and serum obtained in the clinical studies in PROMISE (Task 3.3.2) were analysed for further exploration and validation of functional microbiome studies.

Details on the study design and information on sample collection are described in the "Design clinical protocol with WP4 partners to optimise patient recruitment and sampling for biomarker validation". Briefly, this is a prospective, case-control study conducted across 2 consecutive years in the Wilhelmina Children's Hospital and surrounding general practitioner (GP) offices and general hospitals. Previously healthy infants with different severities of RSV infection (n=74) and healthy infants (n=29), were enrolled in the following sub-groups: (i) healthy infants (n=29); (ii) medically

attended non-hospitalised infants (n=27); (iii) hospitalised, non-ventilated infants (n=24); and (i) ventilated infants (n=23). Stool and serum samples were collected at enrolment (within 96 hours after hospitalisation or start of symptoms), while respiratory tract swabs were collected at the first visit and at 7 weeks ( $\pm 1$  week) after discharge from hospital/emergency department/post onset of symptoms (convalescent time point; RSV-positive infants only). Samples were collected from most of the participants, with few exceptions as indicated in **Figure 1**. *Overall number of samples collected by source (stool, respiratory and serum) for healthy controls and disease groups (medically attended, hospitalised and ventilated infants).* **Figure 1**. Additional information is included in Report D4.4.



**Figure 1.** Overall number of samples collected by source (stool, respiratory and serum) for healthy controls and disease groups (medically attended, hospitalised and ventilated infants). Figure created with BioRender.

Shotgun metagenomic sequencing was performed on stool samples collected at diagnosis, which allowed detailed taxonomic and functional characterisation of the gut microbiota communities (gene prediction and metabolic pathway annotation) associated with RSV severity. 16S rRNA gene amplicon sequencing was performed on nasopharyngeal samples collected at both diagnosis and convalescence timepoints and was used to determine the overall taxonomic composition of the respiratory microbiome associated with RSV severity and convalescence. Additionally, untargeted metabolomics using ultra-high performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) was performed on the serum samples collected at diagnosis, which allowed the determination of the overall metabolic changes upon RSV infection and those associated with increasing RSV severity.

Further integration of microbiome and serum metabolomic information will help identify specific links between microbial groups or metabolites associated with RSV and disease severity, which ultimately could unveil underlying functional mechanisms associated with RSV infection outcomes.

### 1.3. Objectives and key questions

The main goal of this exploratory study is to confirm the most relevant RESCEU biomarkers associated with RSV infection severity by performing functional analyses of the respiratory and gut microbiomes and sera metabolomes. The specific objectives are:

- To validate biomarkers (microbial species and encoded functions) from gut and respiratory

microbiomes that are associated with RSV infection severity and respiratory sequelae.

- To identify additional microbial taxonomical groups and/or functions associated with RSV disease severity.
- To identify potential functional mechanisms associated with RSV infection through integrative analysis of microbiome profiles and serum metabolome.
- To determine respiratory microbiota signatures associated with convalescence status or clinical outcomes (7 weeks after discharge).

## 2. Methods

### 2.1. Data generation

The detailed methods related to data generation are described in D4.4 “Generation of metagenomics and metabolomics raw data”. A brief summary of each method is given below.

#### 2.1.1. Shotgun-metagenomic sequencing data

DNA extraction from stool samples, library preparation and shotgun sequencing were performed by an external vendor. Briefly, DNA was extracted using the Macherey & Nagel Nucleomag DNA Microbiome Kit. The extraction kit was used on a KingFisher extraction robot (Thermo Fisher) and included a bead beating step. This was followed by standard genomic library preparation, QC from extracted DNA and whole-metagenome shotgun sequencing (Illumina NovaSeq 6000 platform with 150 base pairs, paired-end reads). A total of 10 million read pairs (20 million reads) per sample ( $\pm 3\%$ ) was achieved. As mentioned in D4.4, the generated data were of sufficient quality and were transferred to GSK.

#### 2.1.2. 16S rRNA gene amplicon sequencing data

16S rRNA gene amplicon sequencing from nasopharyngeal swabs was performed by an external vendor. In brief, DNA was extracted using the Macherey & Nagel Nucleomag DNA Microbiome Kit. The extraction kit was used on a KingFisher extraction robot (Thermo Fisher) and included a bead-beating step. One 16S V3-V4 amplicon per sample was generated using a two-step PCR protocol. DNA amplicon sequencing on the generated libraries (2 runs) was performed on an Illumina MiSeq platform with paired-end sequencing kit (2x300 bp). A minimum yield of 60,000 read pairs per sample was obtained. As mentioned in D4.4, the generated data were of sufficient quality and were transferred to GSK.

#### 2.1.3. Metabolomics profiling data

Untargeted metabolomics of sera samples was performed by an external vendor. Following sample preparation and quality control, samples were subjected to UPLC-MS/MS to detect, identify and quantify a panel of 5,400 metabolites across 70 major biochemical pathways. Next, data extraction and peak-identification, data QC, compound identification and quantification and curation were performed using vendor’s hardware and software. The resulting data tables were transferred to GSK, including the inferred metabolites, metabolic pathways, chemical annotations, raw data, batch normalised data, imputed data, in addition to some preliminary statistical tests.

#### 2.1.4. Participants and samples associated metadata

Samples and related patients’ metadata were collected as part of WP3. Briefly, samples were received from Universitair Medisch Centrum Utrecht (UMCU). Associated metadata were ingested, curated and formatted for downstream analyses. On top of the information related to the study design, efforts were made to curate any available metadata variables that could confound the respiratory and stool microbiome as well as serum metabolome profiles [8], [9].

In particular, the following metadata types were collected at enrolment and convalescence:

- Age (n=103), sex (n=103), weight (n=86).
- RSV season (n=103), use of respiratory support (n=103), type of respiratory support (oxygen, low-flow nasal cannula, CPAP\_BiPAP, high-flow nasal cannula) (n=103), wheezing (n=103).
- Respiratory rate (n=35), heart rate (n=34), body temperature (n=32), oxygen saturation levels (n=35).
- Presence of concomitant bacterial or viral infection (viral, bacterial, both, none) (n=103).
- Antibiotic use (n=62), respiratory medicine use (n=103).
- Breastfeeding (n=94), Smoking (n=24).

Additional collected metadata included duration of respiratory support, specific respiratory or antibiotic treatment used, route of respiratory medicine administration and confirmed viral and/or bacterial pathogen detected. However, because these additional metadata variables were collected for only a few samples (<20), we did not include them in the analyses.

## 2.2. Bioinformatics data preprocessing

### 2.2.1. Shotgun-metagenomic sequencing data

Quality control of the raw metagenomics sequencing data was performed using FastQC (Andrews, 2010). Sequencing adapters and low-quality reads were trimmed and filtered using fastp [10], and host DNA was removed by mapping the trimmed/filtered reads against the GRCh38 reference genome using Bowtie2 [11]. Host-filtered sequence reads were analysed using MetaPhlan4 [12] to profile microbial taxonomic composition and HUMAnN2 [13] to compute microbial pathways. Assigned taxonomies were converted into the Genome Taxonomy Database (GTDB) latest release (r207, March 2022) [14]. Gene families from the UniProt Reference Clusters were also mapped to the microbial pathways from the MetaCyc metabolic pathway database [15].

Samples with low depth of sequencing (less than 5M classified reads) or that contained more than 50% of unclassified reads were removed from further analyses. Furthermore, unclassified taxon at each taxonomic level were pooled into a new taxon (previous taxonomic level) referred to as 'unclassified\_group'.

### 2.2.2. 16S rRNA gene amplicon sequencing data

The demultiplexed, adapter-removed, paired-end sequencing reads provided by the external vendor were processed using the DADA2 pipeline [16]. In brief, after inspection of read quality profiles, sequences were filtered, trimmed and chimeras removed. Sample decontamination was also performed based on the positive and negative controls using the decontam package [17]. Compositional matrices for each taxonomical level were carried out using the Genome Taxonomy Database (GTDB) latest release (r207, March 2022) [14]. Classifications with low confidence at the genus level (<0.8) were pooled in an arbitrary taxon referred to as 'unclassified\_group'. Each sample was randomly rarefied to 10,000 counts per sample and taxa with low relative abundance (<0.0001) were excluded from further analysis.

### 2.2.3. Metabolomics data

Metabolomics data processing is described in detail in D4.4. Briefly, raw data were extracted, peak-identified, and QC-processed using the service provider's hardware and software. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. An

internal library based on authenticated standards that contains the retention time/index, mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules was used for the identification. More than 4,500 commercially available purified standard compounds were used for analysis on all platforms for determination of their analytical characteristics. A variety of curation procedures were performed to ensure that a high-quality data set was made available for statistical analysis and data interpretation. Peaks were quantified as area-under-the-curve detector ion counts.

## 2.3. Statistical analysis

All the statistical analyses were performed in R using a range of packages including *phyloseq* [18], *microbiome* [19], *vegan* [20], *stats* [21], *factoextra* [22], MaAsLin [23], *mixOmics* [24], *imputeLCMD* [25].

### 2.3.1. Correlated metadata and drivers of microbiome variations

Correlation between metadata variables was assessed using Spearman correlations between continuous variables or between continuous and categorical (transformed into factors and subsequently numeric values). In addition, pairwise chi-squared tests were used to evaluate statistical differences in the proportions of categorical variables among groups. Metadata variables that can affect the microbiome and metabolome composition were identified by performing single distance-based redundancy analysis (dbRDA) with the *capscale* function from *vegan* [20], and estimation of cumulative contribution of significant variables was determined using a forward selection model via the *ordiR2step* function. When possible, these variables known to affect microbiome and/or metabolome were accounted for in the analyses. P-values were adjusted for multiple testing correction using the Benjamini–Hochberg method at 0.05 false discovery rate (FDR).

### 2.3.2. Stool and respiratory microbiome-derived features

Microbial abundances and taxonomic information, or pathway abundances were imported as a *phyloseq* object for further manipulations and analysis using the *phyloseq* package in R [18]. Prior to analyses, different transformations were applied on the taxonomic and functional profiles, including relative transformations (rarefied and relative proportion) and compositional transformations (centred log-ratio (CLR)). The specific transformation applied is indicated for each step in detail. To characterise the composition of the respiratory and stool microbiomes and assess changes between groups (healthy and RSV-infected) and/or timepoints (enrolment, convalescence), the computed taxonomic and functional profiles were used to derive the following microbiome features: (i) Species-level (stool) and genus-level (respiratory) alpha diversity indices; (ii) Clustering based on community composition (respiratory); (iii) Community composition dissimilarity or beta diversity (stool, respiratory) and (iv) abundance of taxonomic groups (genus- and/or species-level), and functions (stool). The Shannon's index of alpha diversity was computed on the rarefied abundance profiles using the *phyloseq* and *microbiome* packages in R [18], [19]. Kruskal–Wallis tests with post-hoc Dunn tests were used to test differences in alpha diversity for more than two groups, while a Wilcoxon rank-sum test was applied to test differences between two groups.

Inter-individual variation or community composition dissimilarity (beta diversity) was computed using the *vegan* R package on functional and species-level Bray-Curtis distance on rarefied bacterial counts after CLR transformation. Permutational analysis of variance (PERMANOVA) testing on beta diversity microbiome data was performed to assess differences between the community composition of the

studied groups. Principal Coordinate Analysis (PCoA) coloured/shaped by timepoint/treatment was additionally used to visualise the community composition using *ggplot2* R package [26].

Clustering of samples based on community composition dissimilarity (beta diversity) was performed using the *kmeans* and *hclust* functions of the *stats* R package, respectively, on the rarefied bacterial read counts [21]. The ideal number of clusters or K value was identified using the elbow method using the *fviz\_nbclust* function from the *factoextra* package in R [22].

Relative abundance at the species or genus level were visualised using stacked bar plots, while CLR-transformed abundances of key species were visualised using jitter plots. All plots were generated using the *ggplot2* R package [26].

Differential taxa abundance analysis was performed on CLR-transformed data, using the Microbiome Multivariable Association with Linear Models (MaAsLin) package in R [23]. For each condition tested, the analysis was performed by controlling for the effect of any of the previously identified confounders (section 2.3.1). Here, species with low relative abundance (<0.01) or present in low prevalence (<10% of samples) were excluded from the analysis. Association p-values were adjusted for multiple testing correction using the Benjamini–Hochberg method at 0.05 FDR. Differences in abundance of significant differentially abundant species were visualised using jitter plots with the *ggplot2* R package.

### 2.3.3. Serum metabolome-derived features

The table of metabolite concentrations was analysed in R. For three samples, analysis was performed twice to assess the robustness of the technique (see D4.4.). The average of the two measurements was kept for the rest of the analysis. Data were then median-normalised. Filtering of metabolites presenting few observations was performed. 186 metabolites out of 1276 were excluded since they were present in less than 80% of the samples in all groups. When missing data did not exceed 20% in all groups (controls, non-hospitalised, hospitalised and ventilated), a left-censored missing data imputation method was applied using the *impute.QRILC* function implemented in the *imputeLCMD* R package [25]. All the metabolites discussed in this report met this criterion. The metabolite concentration was then log transformed (natural logarithm) to improve data distribution.

PCA with and without scaling to unit variance and partial least square discriminant analysis (PLS-DA) were respectively performed using the functions *pca* and *plsda* of the *mixOmics* R package [24]. PERMANOVA was used (*adonis* function of the *vegan* R package [20]) to determine the percentage of variance explained by the group in the dataset. K-means clustering was performed using the *kmeans* function of the *stats* R package [21]. An optimal of three clusters was identified using the elbow method both for within cluster sum of squares and average silhouette performed with the *fviz\_nbclust* function of the *factoextra* R package.

Significantly affected metabolites were identified by a Welch t-test between controls and RSV-positive patients and using ANOVA with Tukey's post-test for comparisons between all groups and RSV-positive groups. The p-value was adjusted for multiple testing using the Benjamini–Hochberg method at 0.05 FDR. Visualisations of metabolite changes were performed in R using the *ggplot2* package [26]. GraphPad Prism v10.1 was used for visualisations.

### 2.3.4. Correlation between microbiome and metabolome

A Spearman's correlation test was performed to identify any significant correlations between metabolites levels and microbial abundance (species-level, genus-level) or pathway abundance, using the *associate* function of the *microbiome* R package [19]. Microbial species/genera were filtered

to remove species with low detection ( $<0.01$ ) or low prevalence ( $<1\%$ ). Abundance values were CLR-transformed *prior* to the analysis. Metabolites were filtered to remove unknown metabolites, and log-transformed prior to analysis. Associated p-values of the correlation tests were adjusted for multiple testing using the Benjamini–Hochberg method at 0.05 FDR. Results were visualised as heatmaps, generated using the *pheatmap* package in R [27], showing only correlation pairs where at least one member had at least one significant correlation (FDR  $<0.05$ ). For visualisations, significant results were further filtered based on the Spearman’s correlation coefficient (e.g.,  $>|0.3|$ ).

## 2.4. Data retention and availability

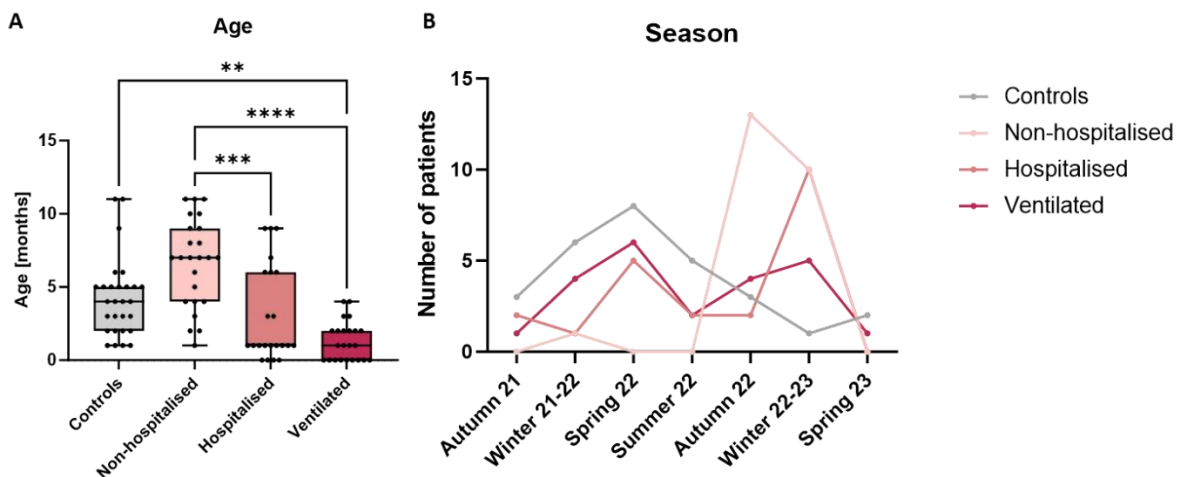
The sequencing data will be deposited in the NCBI SRA repository and the processed metabolomics data will be included in the supplementary materials of the subsequent publication. The scripts used for data analysis will be shared on Github.

## 3. Results

### 3.1. Description of metadata variables

#### 3.1.1. Baseline differences in metadata distribution between groups

First, we investigated whether any baseline differences existed between the four health and disease groups. We identified significant differences in age between the groups (Kruskal-Wallis test,  $p=0.0001$ ; **Figure 2A**). Hospitalised (ventilated and non-ventilated) infants were the youngest (median age=1 month), followed by healthy controls (median age=4 months) and non-hospitalised RSV infants (median age=7 months). We also observed significant differences in seasonality between groups (Fisher's test,  $p=4.9 \times 10^{-4}$ ; **Figure 2B**), with most control samples being taken in the spring of 2022, while most samples from non-hospitalised RSV patients were collected between autumn and winter 2022-2023. These baseline differences could impact the analysis and therefore be considered when interpreting biomarkers of disease and RSV severity.



**Figure 2. Differences in age and seasonality between samples at baseline. (a)** Age (months) differences between RSV (Non-hospitalised, Hospitalised, Ventilated) and control groups at baseline. Boxplots represent the 25th and 75th percentiles (lower and upper boundaries of boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25th and 75th percentiles; whiskers). Statistical significance was assessed using Kruskal-Wallis with Dunn's post-hoc test. **(b)** Line plot showing the number of patients in RSV (Non-hospitalised, Hospitalised, Ventilated) and Control groups for the different seasons (Autumn, Winter, Spring, Summer) over the study period (2021-2023). Significant differences in the number of patients between groups were assessed using a Fisher's exact test. Hospitalised and hospitalised.

#### 3.1.2. Correlated metadata variables

Next, we investigated potential correlations between these variables and found that some of them were not independent from each other. We identified many significant correlations, such as season correlating with age ( $p=1.6 \times 10^{-2}$ ), breastfeeding ( $P=1.0 \times 10^{-3}$ ), antibiotic use ( $P=3.0 \times 10^{-3}$ ) and respiratory medicine use ( $P=1.4 \times 10^{-3}$ ). Furthermore, RSV disease status correlated with sex ( $P=7.0 \times 10^{-3}$ ), respiratory support ( $P=1.0 \times 10^{-3}$ ), antibiotics use ( $P=2.8 \times 10^{-2}$ ) and respiratory medicine use ( $P=4.3 \times 10^{-2}$ ). Wheezing was also correlated with antibiotics use ( $P=1.0 \times 10^{-3}$ ). Finally, presence

of a secondary bacterial or viral infection (“infection status”) correlated with antibiotics use ( $P=0.019$ ) and respiratory medicine use ( $P=3.0 \times 10^{-3}$ ).

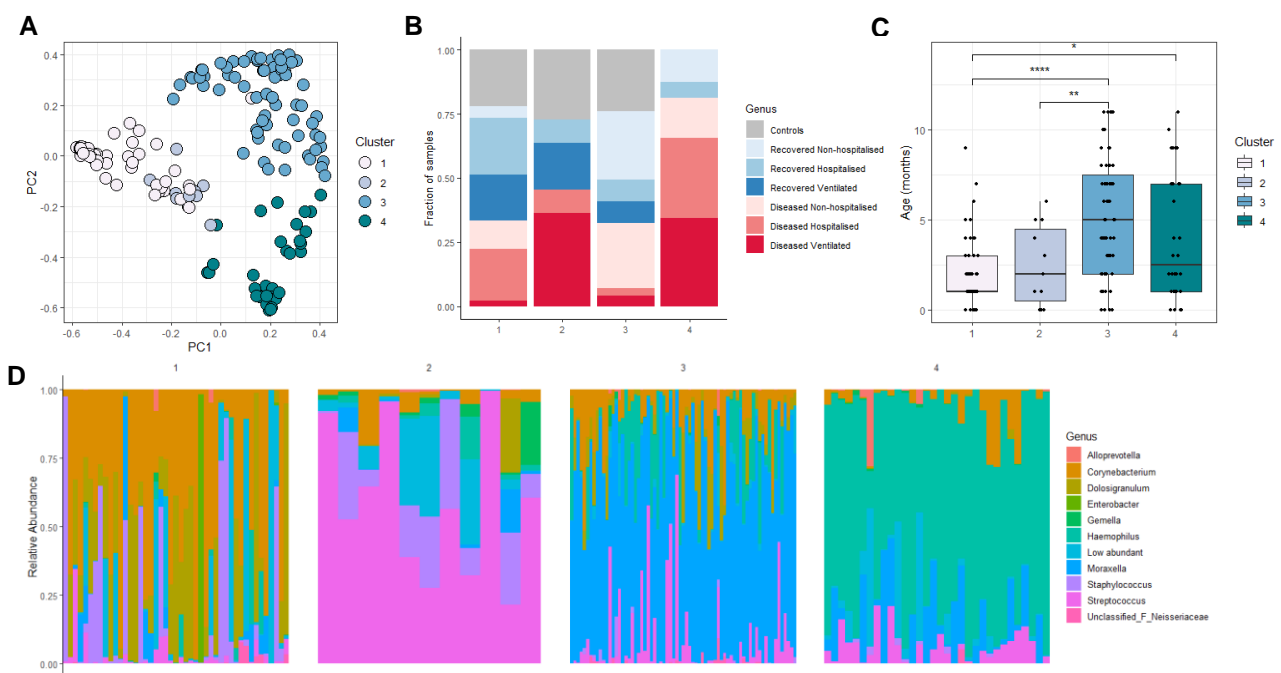
## 3.2. Respiratory microbiome

### 3.2.1. Samples information

Following bioinformatic pre-processing and QC, 159 nasopharyngeal samples of RSV-infected infants and healthy controls were considered for further analysis. Removal of taxa with low relative abundance ( $<0.001$ ) retained 265 genera, which were reduced to 257 after rarefaction.

### 3.2.2. RSV patients clustering at diagnosis

First, we investigated whether differences in the microbiome composition among the four disease severity groups existed. A clustering of all respiratory samples using a K-means approach stratified the microbiome variation into four main clusters (**Figure 3A**). These were reduced to 3 clusters when the analysis was computed on the samples at diagnosis alone (**Annex I, Figure 1**).



**Figure**

**3. Patient clustering based on genus-level respiratory microbiome profiles.** **A)** Principal Coordinate Analysis based on Bray–Curtis dissimilarities showing the nasopharyngeal microbiota variation in RSV patients and controls. Each point represents a nasopharyngeal sample coloured by cluster (1,2,3,4). Clustering of samples was performed on the community composition dissimilarity (beta diversity) using the k-means and hclust R functions. **B)** Stacked bar plot indicating the fraction of RSV and control samples belonging to each of the identified clusters. **C)** Age (months) distribution among the identified clusters, box plots represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles (lower and upper boundaries of boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25<sup>th</sup> and 75<sup>th</sup> percentiles). Statistical significance was assessed using a Wilcoxon’s test. **D)** Relative abundance of ten most abundant genera in the respiratory microbiome for each of the identified clusters, with lower abundant genera included in the group “low abundant” (light blue).

The four clusters were defined by different microbiome composition, each of them being enriched in one or few major genera (**Figure 3D**): Cluster 1 (n=45) included samples characterised by a higher abundance of *Corynebacterium* and *Dolosigranulum*; Cluster 2 (n=11) was enriched by *Streptococcus*; Cluster 3 (n=71) presented higher levels of *Moraxella* and cluster 4 (n=32) was driven by *Haemophilus*. We observed that controls were unevenly distributed across Cluster 1 (n=10), 2 (n=3), and 3 (n=16) but were absent in Cluster 4 (**Figure 3B**). As for RSV patients, Cluster 4 included the most severe RSV cases, in particular high proportions of samples from Ventilated and Hospitalised RSV cases, and none of the recovered subjects. Cluster 3 included the least severe subjects, including Non-Hospitalised RSV and Recovered Non-hospitalised. Cluster 2 included a high proportion of Ventilated RSV and Recovered Ventilated (**Figure 3B**). Finally, Cluster 1 included high proportions of Hospitalised RSV and Recovered Hospitalised (**Figure 3B**). We identified significant age differences between clusters, with Cluster 3 including samples from older infants (**Figure 3C**).

### 3.2.3. Drivers of microbiome variation

Next, we investigated whether any variables had a significant effect on the overall respiratory microbiome composition. We identified that age ( $R^2=0.034$ ,  $P=2 \times 10^{-4}$ ), RSV disease status ( $R^2=0.056$ ,  $P=0.0001$ ), season ( $R^2=0.077$ ,  $P=3 \times 10^{-3}$ ), and breastfeeding ( $R^2=0.087$ ,  $P=7.1 \times 10^{-3}$ ) were significant drivers of variation in the overall cohort (RSV and controls) (**Figure 3A**). Sequencing batch also had a significant effect, although it had the smallest contribution ( $R^2=0.093$ ,  $P=7.5 \times 10^{-3}$ ). Additionally, we found that the presence of a secondary bacterial or viral infection (infection status) ( $R^2=0.071$ ,  $P=1.3 \times 10^{-3}$ ) had the biggest effect on the respiratory microbiome composition in RSV patients (**Table 1B**). Of note, use of antibiotics and respiratory medicines did not represent significant drivers of respiratory microbiome variation after multiple testing correction ( $P=2.8 \times 10^{-2}$ , and  $P=3.5 \times 10^{-2}$ , respectively). Yet, as these variables were significantly correlated with both season and disease status, their impact on the microbiome could be in part reflected by the variation driven by the latter. Considering their significant effect in driving differences in the respiratory microbiome, age, season, breastfeeding, sequencing batch and infection status were considered as confounders, and included in the statistical analysis when assessing specific microbiome-based biomarkers of RSV and disease severity.

**Table 1. Drivers of microbiome variation.** Metadata variables contributing to variations in the respiratory microbiome composition of RSV infants and healthy controls (A) or RSV infants only (B). Variables were identified by performing single dbRDA analysis with the capscale function from vegan, and cumulative contribution of significant variables by using a forward selection model via the ordiR2step function.

#### A) RSV and healthy controls

Variable	R <sup>2</sup> adjusted	Df	AIC	F	Pr(>F)
Age	0.034	11	844	1.5	0.0002
Disease status	0.056	2	843	2.5	0.0001
Season	0.077	6	845	1.5	0.0030
Breastfeeding	0.087	3	846	1.5	0.0071
Sequencing batch	0.093	1	846	1.8	0.0075

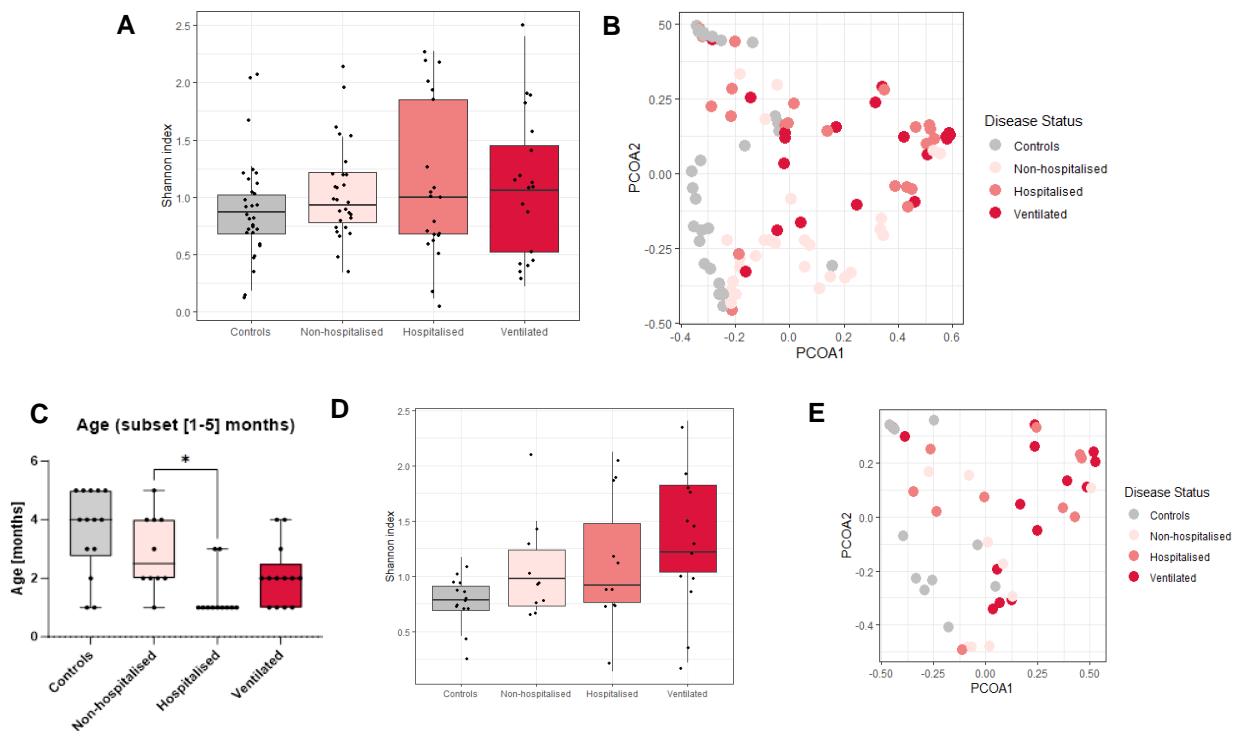
#### B) RSV only

Variable	R <sup>2</sup> adjusted	Df	AIC	F	Pr(>F)
Infection status	0.071	4	385	2.2	0.0013
Visit	0.100	1	384	3.0	0.0002
Season	0.125	5	386	1.4	0.0166
Age	0.158	7	389	1.3	0.0222

### 3.2.4. Microbiota diversity and composition of RSV patients and controls at diagnosis

To assess whether changes in the early-life respiratory microbiome in RSV patients compared to

controls existed, we calculated the microbial richness (Shannon's index of alpha diversity) at genus-level. No differences in richness of respiratory microbiome were observed between groups based on disease status (Kruskal-Wallis rank sum test,  $P=4.3 \times 10^{-1}$ ; **Figure 4A**). An assessment of the overall microbiota community composition at diagnosis (Visit 1) revealed differences in community compositions based on disease status (PERMANOVA,  $R^2=0.1795$ ,  $P=1 \times 10^{-4}$ ). The respiratory microbiome in RSV patients shifted significantly compared to healthy controls (**Figure 4B**), with Ventilated and Hospitalised RSV infants showing increased changes in the community compositions compared to less severe cases (Non-hospitalised RSV, PCoA1 axis). Nevertheless, age also had an impact on the respiratory microbiome composition at diagnosis (PERMANOVA,  $R^2=1.6635$ ,  $P=3 \times 10^{-4}$ ) (**Annex 1, Figure 2**). Given the significant confounding effect of age, we repeated this analysis with infants between 1 and 5 months old alone, while ensuring similar number of observations between groups (Controls=11, Hospitalised=11, Non-hospitalised=10, Ventilated=13). Within this subset of infants, we managed to eliminate most differences in median age between groups, although a small significant difference remained between hospitalised and non-hospitalised RSV infants (**Figure 4C**).



**Figure 4. Richness and microbiome composition in RSV and healthy controls at diagnosis. A)** Genus-level Shannon diversity between RSV patients (red) and control (grey) groups at baseline. Statistical significance was assessed using Kruskal-Wallis with a Dunn's post-hoc test. **B)** Principal coordinate analysis based on Bray–Curtis dissimilarities showing the nasopharyngeal microbiota composition in RSV patients and controls at diagnosis. Each data point represents a nasopharyngeal sample at baseline (V1) coloured by disease status (Controls, Diseased Non-hospitalised, Diseased Hospitalised, Diseased Ventilated). Similar analysis for age subset (1-5 months) (**C, D**).

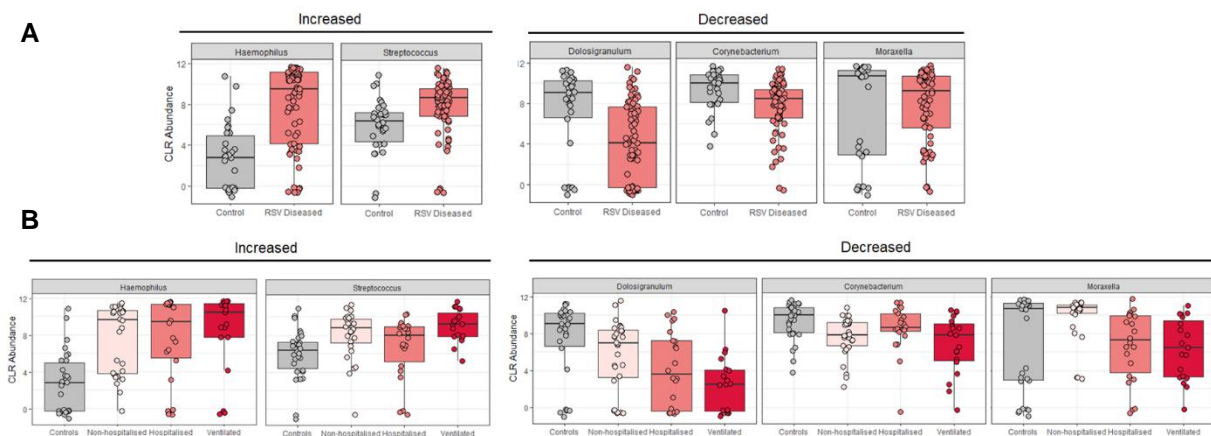
Accounting for the effect of age, we observed a significant difference in microbial richness between disease groups (Kruskal-Wallis rank sum test,  $P=5.1 \times 10^{-2}$ ), with increasing richness alongside the RSV severity axis (**Figure 4D**). In particular, significant differences were driven by differences between the Ventilated RSV group and healthy controls (Dunn's post-hoc test,  $P=3.3 \times 10^{-2}$ ) (**Figure 4D**). Additionally, a significant effect of disease status was observed on the overall respiratory microbiome composition (PERMANOVA,  $R^2=0.1731$ ,  $P=8 \times 10^{-4}$ ), with a clear separation between RSV

patients and healthy controls, as well as between more severe RSV infants (Ventilated and Hospitalised) and milder infections (Non-Hospitalised RSV samples; **Figure 4E**).

Overall, these results confirm that, when removing the confounding effect of age, a shift in both richness and composition of the respiratory microbiome can be observed in RSV groups compared to healthy controls, which is mainly driven by the more severe group (RSV Ventilated).

### 3.2.5. Microbial genera associated with RSV disease and severity

Next, we sought to identify bacterial taxa associated with health and RSV severity groups. For this goal, we performed differential abundance analyses, accounting for the significant variables impacting the microbiome variation (**Tables 1A, 1B**). An overall comparison between RSV patients and healthy infants revealed that diseased groups presented a significant increase in *Haemophilus* and *Streptococcus* genera (**Figure 5A, Annex I Table 1**), while *Corynebacterium* and *Dolosigranulum* were significantly decreased (**Figure 5A; Annex I Table 1**). Conversely, a comparison between RSV severity groups only identified the genus *Moraxella* to be significantly decreased in more severe RSV patients (Hospitalised, Ventilated) (**Figure 5B & Annex I, Table 2**). Notably, among the species associated with RSV, *Haemophilus* and *Dolosigranulum* abundances presented increasing and decreasing trends with RSV disease severity, respectively.

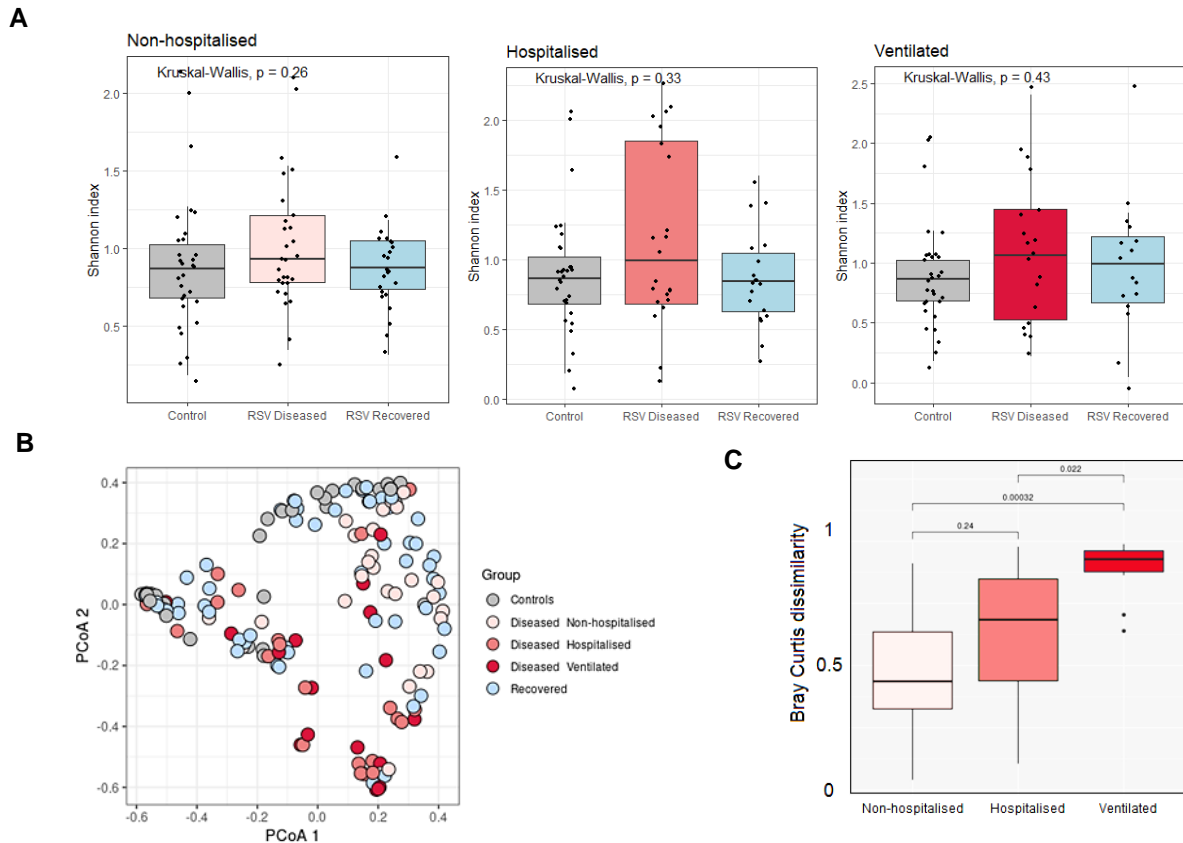


**Figure 5. Respiratory genera shifts in RSV disease at diagnosis compared to healthy controls.** A) CLR-transformed abundance of microbial genera in the respiratory microbiome of RSV and control samples at diagnosis. Box plots represent the 25th and 75th percentiles (lower and upper boundaries of boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25th and 75th percentiles; whiskers). Statistical significance was assessed using Maaslin2 ( $P < 0.05$ ). B) CLR-transformed abundance of microbial genera in the respiratory microbiome of RSV groups and control samples at diagnosis split by disease severity. Box plots represent the 25th and 75th percentiles (lower and upper boundaries of boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25th and 75th percentiles; whiskers). Statistical significance was assessed using Maaslin2 ( $P < 0.05$ ).

### 3.2.6. Microbiota composition and diversity in recovered patients

A major goal of this study was to assess the respiratory microbiota composition in convalescent infants (~7 weeks after discharge) compared to RSV-infected and healthy controls. We did not observe significant differences in microbial richness (Shannon's index of alpha diversity) between

infected and recovered for any of the RSV severity groups (Non-hospitalised, Hospitalised, Ventilated). Yet, the microbial richness appears to shift down towards the levels observed in healthy controls (**Figure 6A**). However, the overall respiratory microbiome composition of recovered samples shifted towards the healthy controls' microbiome (**Figure 6B**). These shifts were higher for samples from severe RSV cases ("Ventilated") compared to milder ones ("Hospitalised" and "Non-hospitalised"), as indicated by the Bray-Curtis dissimilarity (**Figure 6C**); but no changes were observed in the Hospitalised group compared to the Non-Hospitalised (Wilcoxon test,  $P=2.4 \times 10^{-1}$ ).

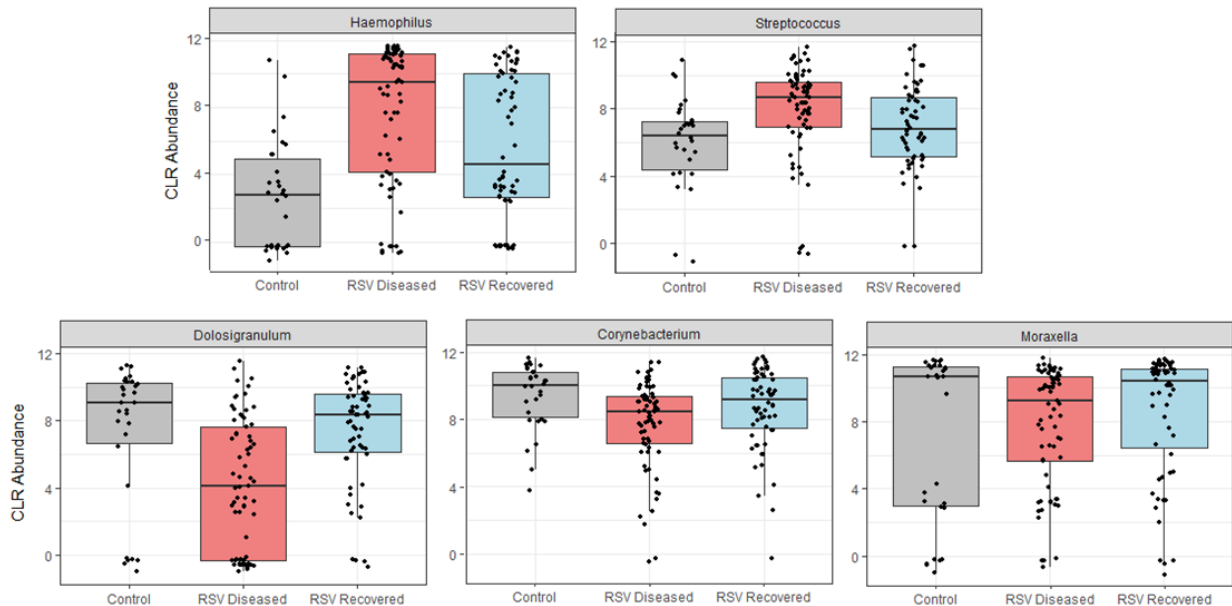


**Figure 6. Respiratory microbiome-derived features in convalescent samples. A)** Genus-level Shannon diversity in each RSV group before and after recovery and compared to controls **B)** PCoA based on Bray–Curtis dissimilarities showing the nasopharyngeal microbiota composition at RSV diagnosis and during RSV recovery. **C)** Bray-Curtis dissimilarity between RSV recovered (Visit 1) and baseline (Visit 2) samples for each RSV disease group. Statistical significance was assessed using Wilcoxon signed-rank test.

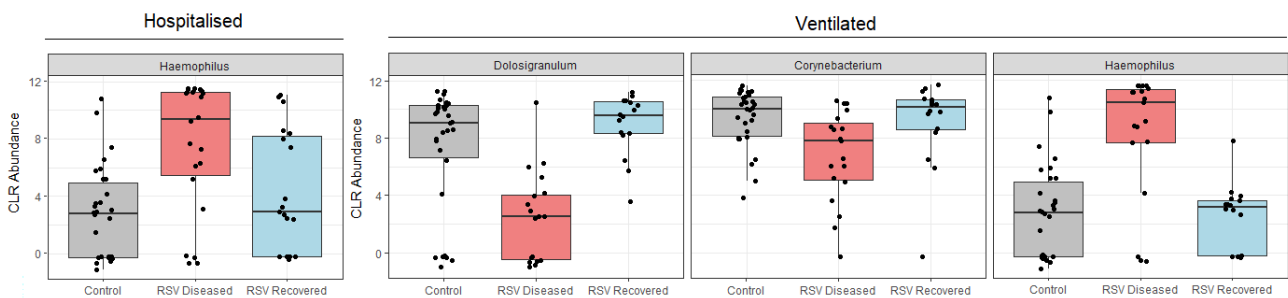
### 3.2.7. Microbial genera changes in recovered individuals

To identify bacterial genera associated with RSV recovery, we performed a differential abundance analysis between convalescent and infected patients for each severity group. Here we also corrected for significant co-variables driving microbiome variation in this group (**Table 1**). In a global comparison of all RSV groups, recovery was generally characterised by decreased levels of *Haemophilus* and *Streptococcus* and increased levels of *Corynebacterium*, *Dolosigranulum* and *Moraxella* (**Figure 7, Annex I Table 3**). Looking into specific differences between RSV severity groups, we observed that recovery changes were higher in the most severe RSV groups (Ventilated, Hospitalised), which was in line with the observed Bray-Curtis dissimilarity described above (**Figure 6C**). In particular, a significant decrease in *Haemophilus* was observed following recovery of the Hospitalised and Ventilated RSV groups (**Figure 8, Annex I Table 4**). Additionally, in the Ventilated group, the

abundances of *Corynebacterium* and *Dolosigranulum* genera significantly increased following recovery (**Figure 8, Annex I Table 4**). Conversely, no differences in respiratory genera were observed for the Non-hospitalised RSV subset following recovery.



**Figure 7. Respiratory genera shifts in RSV disease upon recovery.** CLR-transformed abundance of microbial genera in the respiratory microbiome of control samples and RSV samples at diagnosis (RSV Diseased) and upon recovery (RSV Recovered). Statistical significance between RSV Diseased and RSV Recovered group was assessed using Maaslin2 ( $P < 0.05$ ).



**Figure 8. Respiratory genera shifts in RSV disease upon recovery based on RSV severity.** CLR-transformed abundance of microbial genera in the respiratory microbiome of control samples and RSV groups at diagnosis (RSV Diseased) and upon recovery (RSV Recovered) split by disease severity. Statistical significance between RSV Diseased and RSV Recovered groups was assessed using Maaslin2 ( $P < 0.05$ ).

### 3.3. Stool microbiome

#### 3.3.1. Samples availability

Following bioinformatic pre-processing and quality control of the stool metagenomic data, 92 samples from RSV infants and healthy controls were retained for further analysis. We identified a total of 539 microbial species among all these samples, reduced to 457 after rarefaction.

### 3.3.2. Drivers of stool microbiome variations

First, we investigated potential metadata variables affecting the stool microbiota composition. We found several factors contributing to microbiota variation, including age ( $R^2=0.0765$ ,  $P=1 \times 10^{-4}$ ), disease status ( $R^2=0.086$ ,  $P=1.4 \times 10^{-2}$ ), and breastfeeding ( $R^2=0.0946$ ,  $P=3.9 \times 10^{-2}$ ) (**Table 2A**). These confounders remained when analysing the RSV-infected infants only (**Table 2B**).

**Table 2.** Metadata variables driving significant changes in the gut microbiome composition of RSV infants and healthy controls (A) or RSV infants alone (B). Variables were identified by performing single dbRDA analysis with the capscale function from vegan, and cumulative contribution of significant variables by using a forward selection model via the ordiR2step function.

#### A) RSV and healthy controls

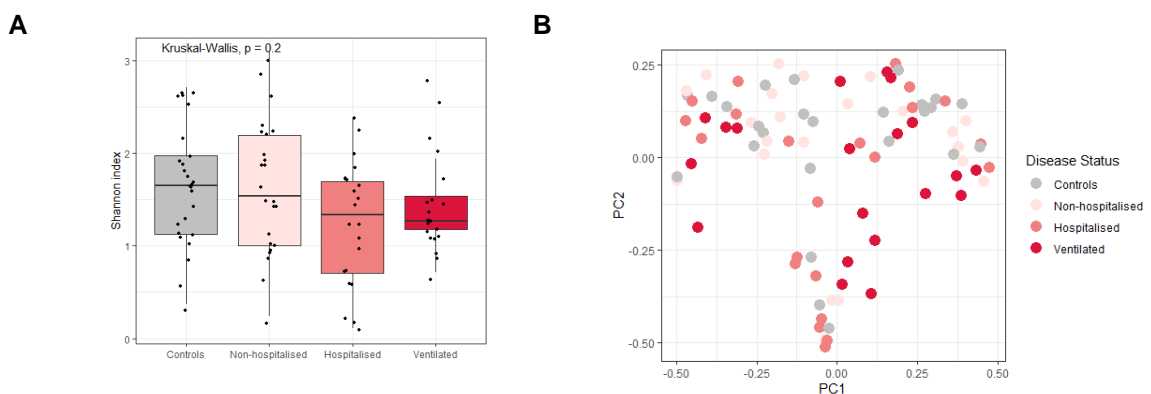
Variable	R <sup>2</sup> adjusted	Df	AIC	F	Pr(>F)
Age (months)	0.0765	11	625	1.6	0.0001
Disease status	0.0865	3	627	1.3	0.0136
Breastfeeding	0.0946	3	628	1.2	0.0390

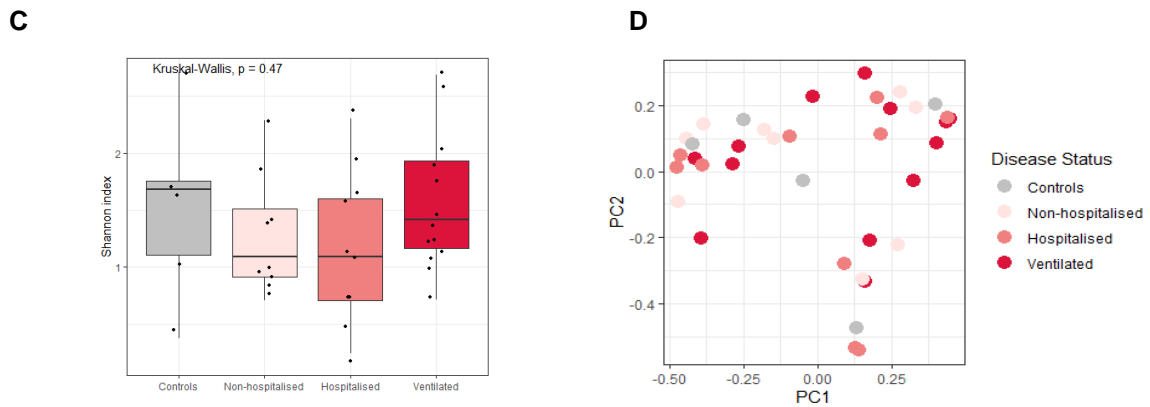
#### B) RSV only

Variable	R <sup>2</sup> adjusted	Df	AIC	F	Pr(>F)
Age (months)	0.1055	10	371	1.6	0.0001
Breastfeeding	0.1186	3	372	1.19	0.0716

### 3.3.3. Microbiota composition and diversity in RSV patients and controls

To assess the impact of RSV infection severity on the stool microbiome compared to healthy controls, we first studied the alpha and beta diversity at the species level. Contrary to the observed effects on the respiratory microbiome, we found no significant differences between RSV samples and controls (Kruskal-Wallis test,  $P=2 \times 10^{-1}$ ; PERMANOVA,  $R^2=0.04672$ ,  $P=6.5 \times 10^{-2}$ ; **Figure 9A**, **Figure 9B**). Given the significant confounding effect of age, here again we subsampled the dataset for samples between 1 and 5 months of age while maintaining a similar number of observations across disease groups (Controls=12, Hospitalised=10, Non-hospitalised=9, ventilated=13). Removal of the confounding effect of age did not yield significant changes in stool microbial richness (Kruskal Wallis test,  $P=4.7 \times 10^{-1}$ ), suggesting that RSV disease has little impact on stool microbial richness (**Figure 9C**). No changes were observed in the community composition between RSV and control groups either (PERMANOVA,  $R^2=0.06715$ ,  $P=7.5 \times 10^{-1}$ ; **Figure 9D**). Overall, these results indicate age is the main driver of differences in stool microbiome between groups, which in turn is affected by breastfeeding, while RSV disease has a minimal impact on the overall richness or microbiome community composition.

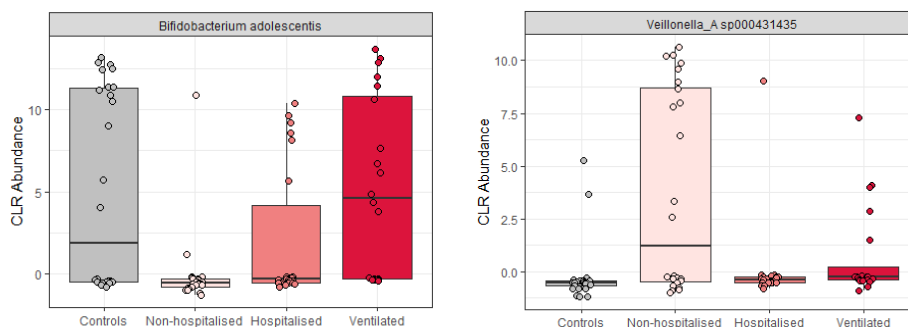




**Figure 9. Stool microbiome-derived features in RSV and healthy controls.** **A, C)** Species-level Shannon diversity between RSV patients (red) and control (grey) groups. Statistical significance was assessed using Kruskal-Wallis with a Dunn's post-hoc test. **B, D)** PCoA based on Bray–Curtis dissimilarities showing the stool microbiota composition in RSV patients and controls coloured by disease status. **C, D** represent infants between 1 and 5 months old.

### 3.3.4. Microbial species and metabolic pathways shifts in RSV

Next, we investigated whether any specific bacterial species presented different abundance levels in diseased infants vs healthy controls or across the disease severity range. Here, we found only one bacterial species, *Bifidobacterium adolescentis*, was significantly decreased in RSV infants compared to healthy controls (FC=-0.87, q-value= $6.9 \times 10^{-2}$ ). This change was mainly driven by the Non-Hospitalised RSV infants, where *Bifidobacterium adolescentis* abundance was significantly decreased compared to healthy controls (FC=-1.09, q= $4.6 \times 10^{-2}$ ; **Figure 10**). The other two severity groups had abundance levels closer to healthy controls, especially the Ventilated RSV group (**Figure 10**). Conversely, *Veillonella A sp000431435* was significantly increased in the Non-Hospitalised RSV group compared to controls (FC=0.3, q= $4.5 \times 10^{-2}$ ), and its abundance remained low for all other RSV groups (**Figure 10**). No other bacterial species presented significant differences in RSV patients.



**Figure 10. Stool microbial species associated with RSV.** Whisker box plots showing the CLR-transformed abundance of top differentially abundant bacteria identified in the different RSV groups and controls. Each dot represent a subject and colour indicate the disease group. Statistical significance was assessed Maaslin2 (p-value < 0.1).

Finally, we used the HMP Unified Metabolic Analysis Network (HUMANn) to profile and quantify microbial metabolic pathways in the RSV and healthy control stool samples. Differential abundance analysis of microbial-derived pathways between RSV and controls revealed very minor differences between these two groups, which were mainly driven by the most severe cases (Ventilated and Hospitalised). We identified only four pathways that were significantly different in RSV; they were related to the mucous layer (peptidoglycan, N-acetylglucosamine, N-acetyl muramate residues) and

(ribo)nucleotides biosynthesis (GMP, isoleucine biosynthesis), which were all downregulated (**Table 3**). No differences were identified when comparing different RSV severity groups against each other.

**Table 3. Stool microbial pathways associated with RSV.** List of microbial-derived pathways differentially abundant between RSV and control groups. Differential abundance analysis was performed using Maaslin2. Only significant results are shown ( $q$ -value  $< 0.1$ ).

Pathway	Comparison	FC	q-value
PWY 7221 (guanosine ribonucleotides biosynthesis)	Ventilated vs Control	-0.004	0.02
PEPTIDOGLYCAN SYN PWY (peptidoglycan biosynth.)	Ventilated vs Control	-0.002	0.06
PWY 6387 (UDP N acetylmuramoyl pentapeptide biosynthesis)	Ventilated vs Control	-0.002	0.06
PWY 3001 (superpathway of L isoleucine biosynthesis)	Hospitalised vs Control	0.002	0.09

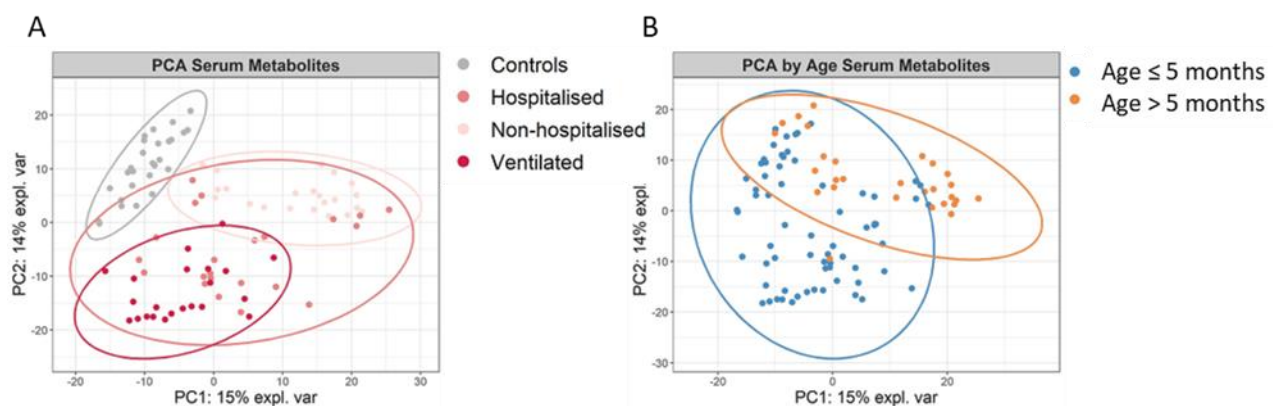
## 3.4. Serum metabolome

### 3.4.1. Sample information

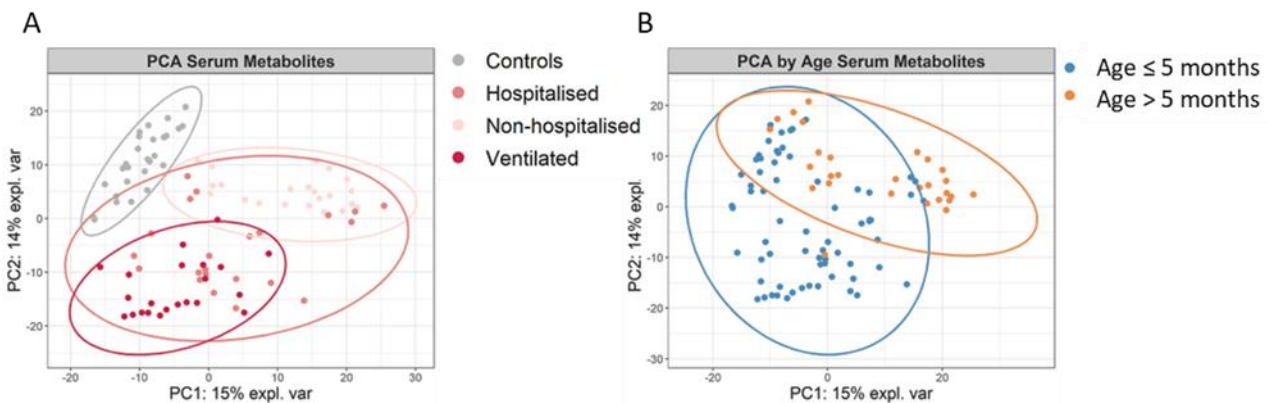
Serum metabolomics results were obtained for 96 independent samples. Following data quality control and filtering of sparsely measured metabolites, a total of 1090 metabolites remained, of which 952 were identified and characterised.

### 3.4.2. Drivers of serum metabolome variation

A PCA on the full metabolome dataset allowed us to identify a main difference between RSV-positive samples and controls ( $R^2=0.2167$ ,  $P=1 \times 10^{-3}$ ). Samples from hospitalised patients presented a higher heterogeneity compared to those from non-hospitalised and ventilated infants ( $P=1 \times 10^{-3}$ ) (**Figure 11A**). The clustering effect by age observed in the microbiome was also identified here, with samples from younger patients ( $\leq 5$  months) and older patients ( $> 5$  months) correlating with the PCA2 axis (



**Figure 11B**). A PCA using the unit variance scaling on all the metabolites (**ANNEX III Figure 1**) showed an increased overlap of the different groups, indicating that metabolites with higher relative abundance may be more important drivers of the variance.



**Figure 11. PCA on serum metabolites. A)** PCA without unit variance scaling. Dots are coloured by groups: controls in grey and non-hospitalised, hospitalised and ventilated from light to dark pink. **B)** Dots are coloured by age: infants ≤5 months in blue and >5 months in orange.

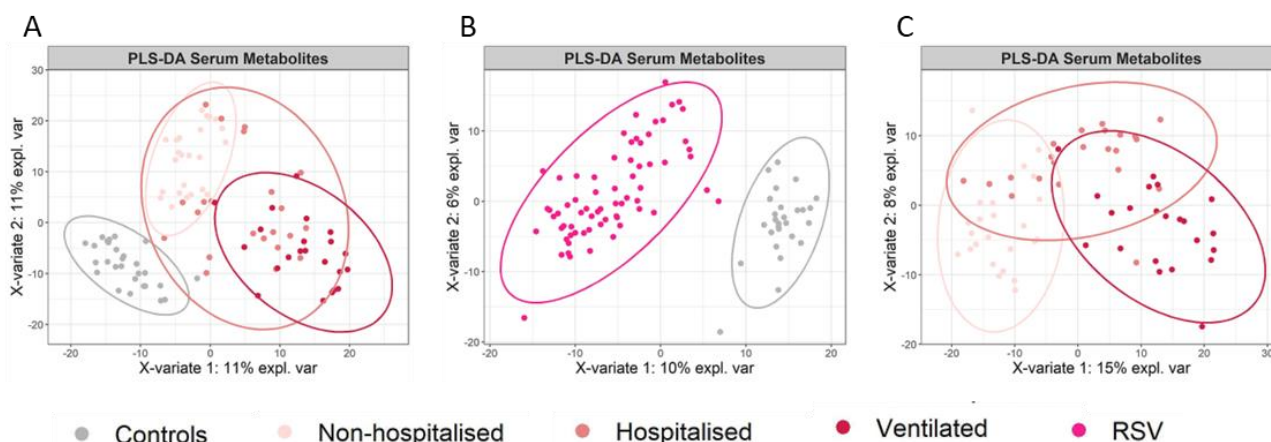
Similar to the respiratory and stool microbiome analysis, a covariate analysis confirmed the significant impact of age on the serum metabolome variation, followed by disease status and breastfeeding (**Table 4**).

**Table 4. Statistically significant confounders of serum metabolome.** Variables were identified by performing single dbRDA analysis with the capscale function from vegan, and cumulative contribution of significant variables by using a forward selection model via the ordiR2step function.

Variable	R <sup>2</sup> adjusted	Df	AIC	F	Pr(>F)
<b>Age</b>	0.256	1	563	7.8	0.0001
<b>Disease status</b>	0.195	3	569	8.0	0.0001
<b>Breastfeeding</b>	0.308	3	560	3.1	0.0001

### 3.4.3. Metabolites driving RSV-positive patients' differences

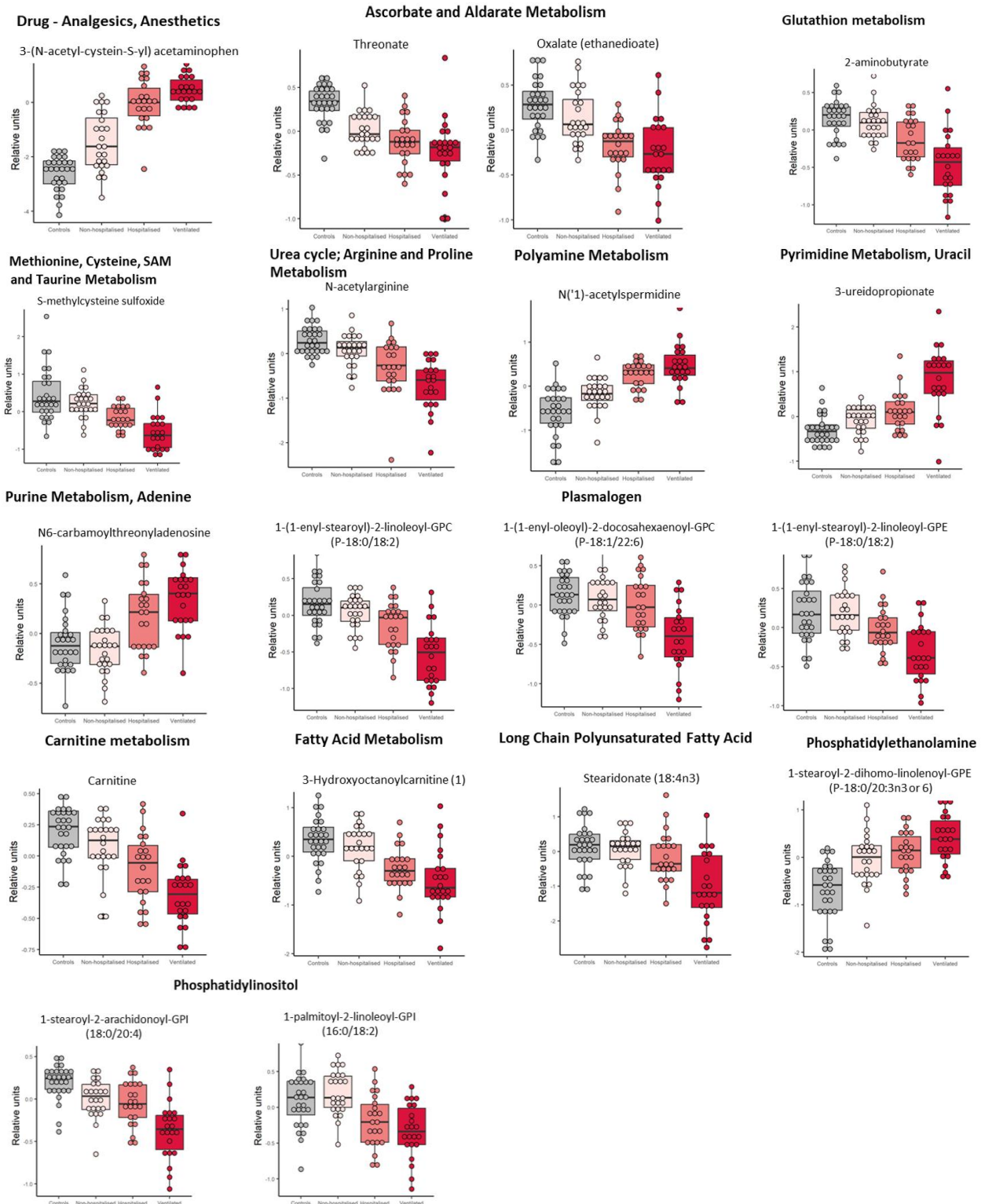
To explore the metabolites driving the separation between the different groups, we used multivariate and univariate statistical analyses. A PLS-DA identified primary metabolites driving differences between RSV infants and controls, and between the four groups (**ANNEX III**; Error! No se encuentra el origen de la referencia. **1, 2, 3**). A clear separation in metabolome profiles was observed between RSV infants and healthy controls (**Figure 12B**). Regarding disease severity, non-hospitalised and ventilated infants displayed two distinct groups overlapping with the hospitalised group, an intermediate stage of disease severity (**Figure 12C**).



**Figure 12. Partial Least Square Discriminant Analysis (PLS-DA) on the serum metabolome. A)** PLS-DA with all samples coloured by groups (controls in grey and RSV-infected in red). **B)** PLS-DA with all samples but only coloured in grey for controls and pink for RSV-positive. **C)** PLS-DA on RSV-positive samples only and coloured from light to dark pink for non-hospitalised, hospitalised and ventilated patients.

Univariate analyses between RSV and controls or between RSV severity groups showed several metabolic changes occurring upon RSV infection (**ANNEX III Table 4**). The main differences between RSV and healthy controls were driven by an enrichment in xenobiotics, primarily derivatives from acetaminophen (2-methoxyacetaminophen sulfate, 3-cystein-S-yl acetaminophen and 4-acetamidophenol; **Figure 13**). These metabolites are probably a result of the medication due to RSV infection rather than a reflection of metabolic changes due to the infection. Similarly, amino acid-related metabolites from various sub-pathways (tyrosine, tryptophan, lactoyl and methionine) were also significantly altered in RSV compared to healthy controls (**Figure 13**). The metabolite vanillic alcohol sulfate, part of tyrosine metabolism, is also used in antibacterial medicines and its increase in RSV might simply reflect the use of antibacterial medication in RSV patients.

A comparison between disease severity groups showed that xenobiotic-derived metabolites were also significantly altered, but the highest variations were observed in lipids (plasmalogens, phosphatidylcholine, phosphatidylinositol) and nucleotide and amino acid metabolism metabolites (**Figure 13**). Several lipids decreased alongside RSV severity, particularly plasmalogens, whereas phosphatidylethanolamines were increased. Additionally, metabolites from the carnitine, fatty acids, and the ascorbate and aldarate metabolic pathways also decreased alongside RSV severity (**Figure 13**). Similar to previous observations, the differences between hospitalised and ventilated or non-hospitalised RSV infants were less pronounced, probably due to the high heterogeneity of the hospitalised RSV group.



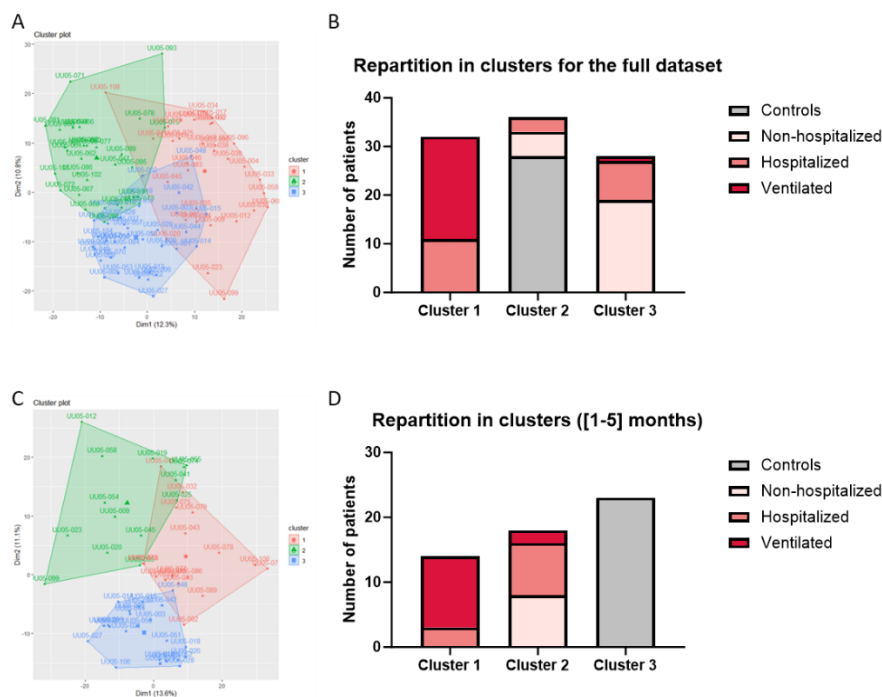
**Figure 13. Altered metabolites between healthy and diseased groups.** These correspond with the xenobiotic, vitamins and cofactors, amino acid and nucleotide super pathways (above) and from the lipids super pathway (below). P-value and q-value from ANOVA <0.05.

Finally, several amino acids belonging to various pathways were found to be decreased in RSV and alongside RSV severity, with the exception of those involved in polyamine metabolism, which were instead increased in RSV (**Figure 13**). Members of the nucleotide metabolism also tended to be increased alongside RSV severity (**Figure 13**).

To account for the confounding effect of age, we also investigated whether these findings were maintained when analysing a patient subset with age between 1-5 months. Overall, when accounting for age, several metabolites' alterations lost significance (e.g., 1-palmitoyl-2-linoleoyl-GPI (phosphatidylinositol) and N6-carbamoyl-threonyl-adenosine (purine)). However, the most pronounced metabolic shifts driving RSV and RSV severity were maintained, suggesting these may be true RSV-related signals and not confounded by baseline age differences.

### 3.4.4. Clustering on serum metabolome

Similar to the respiratory microbiome analysis, we performed a clustering analysis on the metabolomic data, which identified three separate clusters (**Figure 14A-B**). Cluster 1 included the most severe RSV infants (Ventilated, Hospitalised) and was driven by higher abundances of xenobiotics, mainly involved in benzoate metabolism (methyl 4-hydroxybenzoate sulfate) (**Annex III, Table 4**). Cluster 2 included mostly healthy infants and few RSV (Non-hospitalised, Hospitalised) and its composition was driven by amino acids, mainly involved in histidine metabolism (trans-urocanate). Cluster 3 included an intermediate profile, mainly Non-Hospitalised RSV, followed by Hospitalised RSV infants, and its composition was driven by various metabolites, including lipids, amino acids, vitamins and co-factors belonging to different pathways, with no clear common function. A clustering on the subset of patients between 1 and 5 months old also yielded three clusters following the gradient of RSV disease severity (Figure 14C-D).



**Figure 14. K-means clustering on serum metabolome data. A)** Biplot of the three clusters generated on the full dataset. **B)** Repartition of the number of patients of each group in the three clusters generated on the full dataset. **C, D)** Reanalysis using the subset of infants aged 1-5 months.

### 3.5. Microbiome-metabolome associations

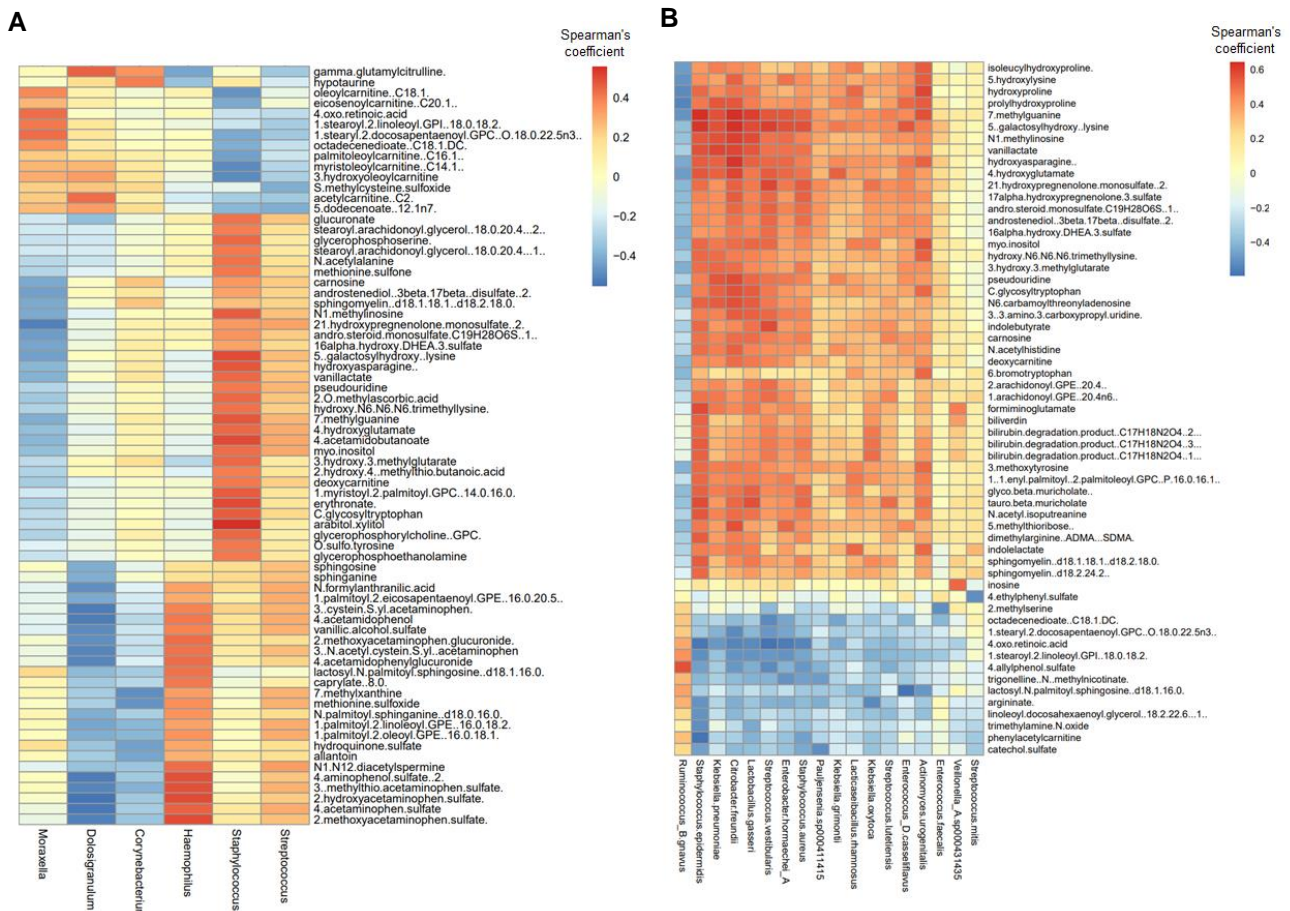
To identify potential mechanistic links between the identified signatures in the stool or respiratory microbiomes and the metabolome, we first explored relationships between the microbiome and metabolic clusters. Overall, we observed extensive overlap between the respiratory microbiome and metabolome clusters (**Annex II, Table 1**), primarily for the most severe samples, with the highest overlap being between respiratory microbiome cluster 4 (*Haemophilus* cluster) and metabolome Cluster 1, both of which included most of the ventilated and hospitalised infants.

Furthermore, we sought for associations between respiratory and stool microbiome components (genera, species, pathways) and serum metabolite levels. We identified several significant positive correlations between the respiratory microbiome and the metabolome (**Figure 15A**), in particular between the health-associated genus *Dolosigranulum* and various fatty acids and fatty acid metabolism metabolites (including phosphatidylcholine, phosphatidylinositol and plasmalogen), which were all found to be decreased in RSV and alongside RSV severity. Conversely, RSV-associated *Haemophilus* and *Streptococcus* genera positively correlated with various xenobiotics, especially derivatives of acetaminophen (**Figure 15B**), which were increased in RSV patients and alongside RSV severity. Finally, *Staphylococcus* and *Streptococcus* positively correlated with various amino acids, phosphatidylcholine, and purine and pentose metabolism metabolites. Interestingly, while amino acids and phosphatidylcholine were decreased in RSV, purine and pentose metabolites were significantly increased.

Several correlations were also identified between gut microbial species and serum metabolites (**Figure 15B**). In particular, *Ruminococcus gnavus* positively correlated with the xenobiotic 4-allylphenol sulfate, phosphatidylinositol and plasmalogen. Additionally, various species (*Citrobacter freundii*, *Staphylococcus epidermidis* and *Staphylococcus aureus*, *Lactobacillus gasseri*, *Streptococcus vestibularis*, *Klebsiella pneumoniae*, *Actinomyces urogenitalis*, and *Enterobacter hormaechei A*) were positively correlated with metabolites involved in amino acid metabolism (alanine and aspartate, glutamate, histidine, lysine, tyrosine, and tryptophan), purine and pyrimidine metabolism (methylguanidine, pseudouridine) and lipid metabolism, such as plasmalogen, androgenic and pregnenolone steroids, and primary bile acid metabolites (glycol beta-muricholate) among others (**Figure 15B**).

Finally, positive correlations between gut microbial-derived pathways and serum metabolites included those between L-histidine degradation (PWY 5030 L histidine degradation III) and phosphatidylinositol (1-stearoyl-2-oleoyl-GPI (18:0/18:1)). Biosynthesis of UDP glucose derived O-antigens pathway (WY 7328 superpathway) was associated with phenylacetylglutamine, and the methanogenesis from acetate pathway with the xenobiotic piperidinone and phenylacetylglutamine (**Annex II, Figure 1**).

Overall, variation in microbial components from the respiratory microbiota corresponded with changes in serum metabolites, but we did not explore causal or directional links. Nevertheless, the combination of respiratory taxa and serum metabolites could lead to defining biomarkers for RSV disease severity.



**Figure 15. Correlations between respiratory and gut microbiome and serum metabolome. A)** Heatmap showing the Spearman's coefficient of the correlation between respiratory genera and serum metabolites. Red indicates a positive correlation, and blue a negative correlation. Correlations were performed using Spearman correlations and Benjamini-Hochberg adjustment for multiple comparison. Significant correlations were identified for those with  $p$ -adjusted  $<0.05$ . For visualisation purposes, only respiratory genera and serum metabolites with at least one significant correlation with Spearman's coefficient  $>|0.4|$  and with  $p$ -value  $<0.05$  are shown. **B)** Heatmap showing the Spearman's coefficient of the correlation between stool microbial species and serum metabolites. Red indicates a positive correlation, and blue a negative correlation. Correlations were performed using Spearman correlation and Benjamini-Hochberg adjustment for multiple comparison. Significant correlations were identified for those with  $p$ -adjusted  $<0.05$ . For visualisation purposes, only bacterial species and serum metabolites with at least one significant correlation with Spearman's coefficient  $>|0.5|$  and  $p$ -value  $<0.01$  are shown.

## 4. Discussion

Here we conducted an integrated analysis of the respiratory and gut microbiomes, along with the serum metabolome of infants to identify biomarkers associated with RSV severity. We identified several co-variables, primarily age, season, and breastfeeding, that influenced variations in the microbiomes and metabolomic profiles (**Table 1**, **Table 2**, **Table 4**). Furthermore, within RSV groups, the presence of a concomitant viral or bacterial infection was also an important driver of respiratory microbiome variation (**Table 1B**). Of note, these variables were not independent of each other, as age correlated with both breastfeeding and season. Given the importance of these confounders in driving variation in both the microbiome and metabolome, future studies should aim to collect as much information about these features as possible to enable for their potential confounding effect to be accounted for when assessing the main study outcomes.

Overall, our study revealed that RSV infection significantly impacts the respiratory microbiome and serum metabolome, while its effect on the gut microbiome is minimal. We observed an increase in richness and shifts in the composition of the respiratory microbiome during RSV infection, with contrasting changes noted in recovered RSV infants, although the levels did not fully reach those of healthy infants (**Figure 4**, **Figure 6**). These changes were most pronounced in the most severe Ventilated RSV infants. Notably, the increase in respiratory microbiome richness with disease severity contradicts what is typically observed in the gut, suggesting an increase in the number of respiratory genera during RSV infection [28].

This work also underscored the importance of patient stratification when searching for biomarkers in clinical settings. We found that RSV patients and healthy controls could be clustered into four groups based on their respiratory microbiome profiles, and three clusters based on the serum metabolome profiles, which considerably overlapped with each other (**Figure 3**, **Figure 14**). Health-associated changes were easier to define, as microbiome clusters that included healthy or recovered RSV infants were characterised by higher abundances of *Corynebacterium* and *Dolosigranulum* (**Figure 3**). Differential abundance analysis also found these genera were significantly decreased in RSV and associated with recovery (**Figure 5**, **Figure 8**). These findings align with previous results from RESCEU, where *Dolosigranulum* and *Corynebacterium* abundance were associated with health. Interestingly, this health-related respiratory cluster overlapped with a serum metabolome cluster whose composition was primarily driven by amino acids involved in histidine metabolism. Interestingly, histidine metabolism has been previously linked to 10-formyl-tetrahydrofolate-dependent purine biosynthesis which is involved in (viral) DNA replication [29]. Conversely, RSV patients were more heterogeneous in both their respiratory microbiome and serum metabolome. In particular, respiratory microbiome clusters included the most severe RSV infants (Cluster 2 and 4) were dominated by higher abundances of *Haemophilus* and *Streptococcus* (**Figure 3**), in line with previous findings from RESCEU. Differential abundance analysis also confirmed that these genera are enriched in RSV and may drive RSV severity (**Figure 5**). Notably, severe RSV clusters overlapped with the serum metabolome cluster driven by xenobiotics (Cluster 1) (**Annex II Table 1**). Correlation analysis confirmed these links, especially with derivatives of acetaminophen, which were also found to be the main metabolic alteration in RSV compared to controls (**Figure 15**, **Figure 13**). These metabolites are derivatives of paracetamol metabolism by the gut microbiota and are reflective of medicine administration rather than a real biomarker of RSV severity. Unfortunately, the clinical information collected did not provide an exhaustive report of medication intake, especially related to acetaminophen/paracetamol. Additionally, *Streptococcus* was found to correlate with phosphatidylcholine (PC), amino acids, and purine and pentose metabolism metabolites (**Figure 15**). Univariate analysis also found these compounds to be increased in RSV, which were mainly increased in RSV and associated with increasing RSV severity (**Figure 13**). Catabolism of amino

acids plays an important role during inflammation, regulating the function of various immune cells such as effector T cells and regulatory T cells [30]. The second respiratory cluster that included the least severe RSV subjects was dominated by *Moraxella* and overlapped with the metabolomic cluster characterised by a mixed profile of metabolites, including lipids, amino acids, vitamins and co-factors, with no clear common pathway (**Annex II Table I**). This heterogeneity could be due to the confounding effect of age, whereby the Non-Hospitalised group included infants with a significantly higher median age compared to all the other groups (**Figure 2A**). Observations concerning *Moraxella* in RESCEU were also ambiguous, as a *Moraxella*-dominated microbial profile was associated with a lower risk of medically attended RSV, yet *Moraxella* abundance was also associated with RSV disease severity. Previous studies have identified an association between higher *Moraxella* abundance and subsequent occurrence of RSV infection [31], and with earlier and more frequent acute respiratory infection [32], [33], [34], [35], [36]. These discrepancies in results might be due to several reasons, one being the different distribution of RSV patients in the two studies, and the confounding effect of age. For instance, RESCEU was conducted mainly with severe RSV infants, while we also considered a high proportion of non-hospitalised RSV cases.

In this study, we could not record any differences in the gut microbiome richness or composition of RSV infants compared to healthy controls (**Figure 9**). These results are in line with findings from RESCEU, where minor differences were observed at the genus level between hospitalised and non-hospitalised RSV, but not compared with healthy controls. When looking at individual changes at the species level accounting for any confounding effects (including age), we only found an increase of *B. adolescentis* and decrease of *Veillonella A sp000431435* in non-hospitalised RSV infants compared to healthy controls (**Figure 10**). Similar to other *Bifidobacterium* species, *B. adolescentis* is an important member of the early-life gut microbiome, and its abundance has been reported to decrease with age (36, 31). Furthermore, *Veillonella* is known to be a key component of the oral microbiota early in life, reaching its highest abundance at 3-6 months of age, to then undergo a decline towards 2-3 years of age [38], [39]. Because non-hospitalised RSV infants were characterised by a significantly higher age compared to the other RSV groups (**Figure 2A**), the observed changes in the stool microbiome may be attributed to an age-related effect, despite this co-variate being used as a correction factor in the multivariate model.

The study of the serum metabolome also revealed several metabolic alterations during RSV infection, mainly in more severe RSV infants (Hospitalised, Ventilated) (**Figure 11, Figure 12**). We found plasmalogens, an important cell membrane antioxidant, decreased alongside RSV severity, which could represent a marker of oxidative stress [40]. A decrease in metabolites from the ascorbate, aldarate and glutathione metabolic pathways in RSV and alongside RSV severity may also indicate increased oxidative stress. Furthermore, the decreased levels of carnitine, fatty acids and long-chain polyunsaturated fatty acids alongside RSV severity could reflect a decrease in energetic metabolism in RSV infants, potentially linked to decreased food intake, although no dietary information was available in this study to confirm this hypothesis. Finally, several metabolites belonging to amino acid metabolic pathways were also altered in RSV patients, but these changes may be more reflective of a general perturbed state in RSV rather than a specific RSV signature. However, since age was found to be an important driver of metabolic variations, future studies with a higher number of patients and of a similar age could validate our findings.

Associating microbiome data with metabolome data can help to better understand the overall functional role of the microbiome on host health and the role of altered host-microbiome interactions during disease. Here, we found several correlations between alterations of serum metabolites and the nasopharyngeal microbiome, while very few associations were identified with gut microbiome components, showing the biggest impact of the former on metabolic alterations during RSV. Among the major shifts observed in serum metabolites in RSV infants and alongside RSV severity, we can

assume that part of these changes, in particular those related to xenobiotics, amino acids, fatty acids and purine and pentose metabolism, may be directly caused by shifts in the respiratory microbiome, suggesting that respiratory microbiota might have a functional role in RSV severity. Nevertheless, further *in vitro* and *in vivo* studies would be required to confirm the role of these microbiome-metabolite interactions on host metabolism during RSV.

Despite our findings, a series of limitations should be considered, primarily the lack of age-matched samples. As we observed, the Ventilated RSV group had a significantly lower age compared to the control group, and significant age differences were also observed between RSV severity groups. Age represented a key confounder and should be taken into account in future studies. Furthermore, we observed that the presence of a concomitant viral or bacterial infection was also a major driver of microbiome variation. This suggests that changes observed in RSV could partially be a consequence of the presence of a secondary infection. Thirdly, we could not identify any biomarker of disease susceptibility, as samples were collected at diagnosis and not prior to infection. Finally, the convalescent period in this study was shorter than the standard period of about 1-3 years, which translated into a not fully healthy microbiota in convalescent infants. Future studies investigating microbiome or metabolic markers of RSV susceptibility or recovery should implement a later sample collection to ensure correct evaluation of convalescence markers.

## 5. Conclusion and next steps

During RSV infection, a clear change in the respiratory microbiome and serum metabolic profiles was observed, while the gut microbiome was less impacted. During recovery, the respiratory microbiome shifted back towards a health-like microbiome. These changes were also identified across the disease severity spectrum defined in this study. We also found associations between specific microbial components and metabolites, which will be further explored in future analysis, aiming to infer mechanistic links. Future studies should be age-matched and include samples of healthy infants before disease diagnosis, which would help in identifying microbiome biomarkers of disease susceptibility and severity.

## 6. Repository for primary data

Raw and processed data are currently stored in GSK servers.

## 7. References

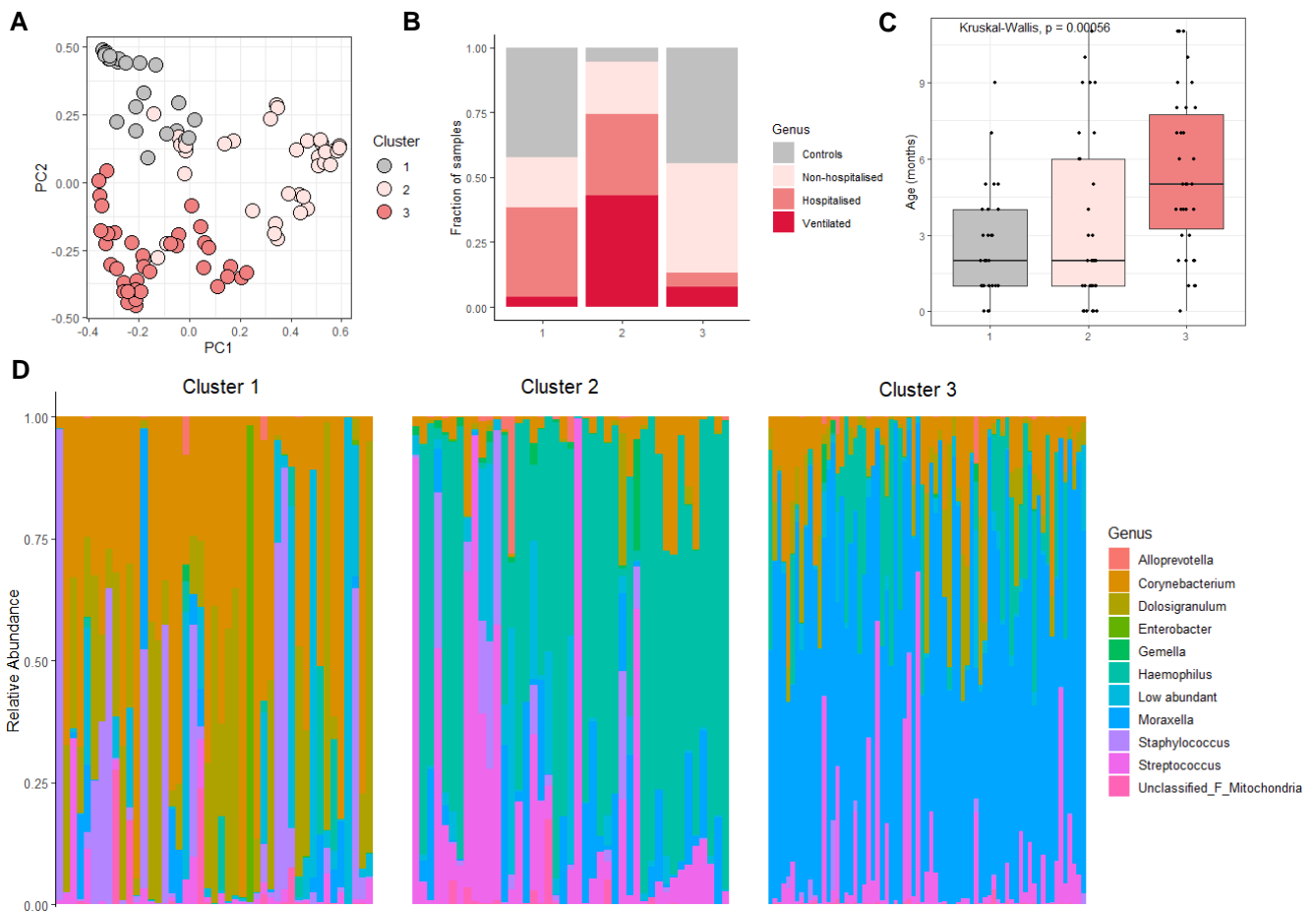
- [1] Y. Li *et al.*, “Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis,” *The Lancet*, vol. 399, no. 10340, pp. 2047–2064, 2022, doi: 10.1016/S0140-6736(22)00478-0.
- [2] P. M. Brown, D. L. Schneeberger, and G. Piedimonte, “Biomarkers of respiratory syncytial virus (RSV) infection: Specific neutrophil and cytokine levels provide increased accuracy in predicting disease severity,” *Paediatr. Respir. Rev.*, vol. 16, no. 4, pp. 232–240, 2015, doi: 10.1016/j.prrv.2015.05.005.
- [3] K. Jefferies *et al.*, “Presumed risk factors and biomarkers for severe respiratory syncytial virus disease and related sequelae: Protocol for an observational multicenter, case-control study from the respiratory syncytial virus consortium in Europe (RESCEU),” *J. Infect. Dis.*, vol. 222, no. Suppl 7, pp. S658–S665, 2021, doi: 10.1093/INFDIS/JIAA239.
- [4] J. G. Wildenbeest *et al.*, “Respiratory syncytial virus consortium in Europe (RESCEU) birth cohort study: Defining the burden of infant respiratory syncytial virus disease in Europe,” *J. Infect. Dis.*, vol. 222, no. Suppl 7, pp. S606–S612, 2020, doi: 10.1093/INFDIS/JIAA310.
- [5] D. Öner *et al.*, “Biomarkers for disease severity in children infected with respiratory syncytial virus: A systematic literature review,” *J. Infect. Dis.*, vol. 222, no. Suppl 7, pp. S648–S657, 2020, doi: 10.1093/INFDIS/JIAA208.
- [6] N. Rifai, M. A. Gillette, and S. A. Carr, “Protein biomarker discovery and validation: The long and uncertain path to clinical utility,” *Nat. Biotechnol.*, vol. 24, no. 8, pp. 971–983, 2006, doi: 10.1038/nbt1235.
- [7] V. Sencio, M. G. Machado, and F. Trottein, “The lung–gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes,” *Mucosal Immunol.*, vol. 14, no. 2, pp. 296–304, 2021, doi: 10.1038/s41385-020-00361-8.
- [8] G. Falony *et al.*, “Population-level analysis of gut microbiome variation,” *Science*, vol. 352, no. 6285, pp. 560–564, 2016, doi: 10.1126/science.aad3503.
- [9] I. Vujkovic-Cvijin, J. Sklar, L. Jiang, L. Natarajan, R. Knight, and Y. Belkaid, “Host variables confound gut microbiota studies of human disease,” *Nature*, vol. 587, no. 7834, pp. 448–454, 2020, doi: 10.1038/s41586-020-2881-9.
- [10] S. Chen, Y. Zhou, Y. Chen, and J. Gu, “fastp : an ultra-fast all-in-one FASTQ preprocessor,” pp. 884–890, 2018, doi: 10.1093/bioinformatics/bty560.
- [11] B. Langmead and S. L. Salzberg, “Fast gapped-read alignment with Bowtie 2,” vol. 9, no. 4, pp. 357–360, 2012, doi: 10.1038/nmeth.1923.
- [12] A. Blanco-Miguez *et al.*, “Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlan 4,” *bioRxiv*, p. 2022.08.22.504593, 2022.
- [13] E. A. Franzosa *et al.*, “Species-level functional profiling of metagenomes and metatranscriptomes,” *Nat. Methods*, vol. 15, no. 11, pp. 962–968, 2018, doi: 10.1038/s41592-018-0176-y.
- [14] P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, and D. H. Parks, “GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database,” *Bioinformatics*, vol. 36, no. 6, pp. 1925–1927, 2020, doi: 10.1093/bioinformatics/btz848.
- [15] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes—a 2019 update,” *Nucleic Acids Res.*, vol. 48, no. D1, pp. D455–D453, 2020, doi: 10.1093/nar/gkz862.
- [16] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes, “DADA2: High-resolution sample inference from Illumina amplicon data,” *Nat. Methods*, vol. 13, no. 7, pp. 581–583, 2016, doi: 10.1038/nmeth.3869.
- [17] N. M. Davis, D. M. Proctor, S. P. Holmes, D. A. Relman, and B. J. Callahan, “Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data,”

- Microbiome*, vol. 6, no. 1, p. 226, Dec. 2018, doi: 10.1186/s40168-018-0605-2.
- [18] P. J. Mcmurdie and S. Holmes, “phyloseq : An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data,” vol. 8, no. 4, 2013, doi: 10.1371/journal.pone.0061217.
- [19] L. Lahti and S. Sudarshan, “Tools for microbiome analysis in R.,” 2017, [Online]. Available: URL: <http://microbiome.github.com/microbiome>
- [20] J. Oksanen *et al.*, *vegan community ecology package version 2.5-7 November 2020*. 2020.
- [21] “R: The R Stats Package.” Accessed: Apr. 19, 2024. [Online]. Available: <https://search.r-project.org/R/refmans/stats/html/00Index.html>
- [22] A. Kassambara, “factoextra : Extract and Visualize the Results of Multivariate Data Analyses.” Accessed: Apr. 19, 2024. [Online]. Available: <https://rpkgs.datanovia.com/factoextra/index.html>
- [23] H. Mallick *et al.*, “Multivariable association discovery in population-scale meta-omics studies,” *PLoS Comput. Biol.*, vol. 17, no. 11, p. e1009442, Nov. 2021, doi: 10.1371/journal.pcbi.1009442.
- [24] F. Rohart, B. Gautier, A. Singh, and K.-A. L. Cao, “mixOmics: An R package for ‘omics feature selection and multiple data integration,” *PLoS Comput. Biol.*, vol. 13, no. 11, p. e1005752, Nov. 2017, doi: 10.1371/journal.pcbi.1005752.
- [25] C. Lazar, T. Burger, and S. Wiczorek, “A Collection of Methods for Left-Censored Missing Data Imputation.” CRAN, Jun. 10, 2022. Accessed: Apr. 18, 2024. [Online]. Available: <http://cran.nexr.com/web/packages/imputeLCMD/imputeLCMD.pdf>
- [26] H. Wickham, *ggplot2*. in *Use R!* Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-24277-4.
- [27] R. Kolde, *Implementation of heatmaps that offers more control over dimensions and appearance*. 2019. Accessed: Apr. 19, 2024. [Online]. Available: <https://cran.r-project.org/web/packages/pheatmap/index.html>
- [28] T. P. Wypych, L. C. Wickramasinghe, and B. J. Marsland, “The influence of the microbiome on respiratory health,” *Nat. Immunol.*, vol. 20, no. 10, pp. 1279–1290, Oct. 2019, doi: 10.1038/s41590-019-0451-9.
- [29] H. Delattre, K. Sasidharan, and O. S. Soyer, “Inhibiting the reproduction of SARS-CoV-2 through perturbations in human lung cell metabolic network,” *Life Sci. Alliance*, vol. 4, no. 1, p. e202000869, Nov. 2020, doi: 10.26508/lsa.202000869.
- [30] Z.-N. Ling, Y.-F. Jiang, J.-N. Ru, J.-H. Lu, B. Ding, and J. Wu, “Amino acid metabolism in health and disease,” *Signal Transduct. Target. Ther.*, vol. 8, no. 1, pp. 1–32, Sep. 2023, doi: 10.1038/s41392-023-01569-3.
- [31] A. Grier *et al.*, “Temporal Dysbiosis of Infant Nasal Microbiota Relative to Respiratory Syncytial Virus Infection,” *J. Infect. Dis.*, vol. 223, no. 9, pp. 1650–1658, May 2021, doi: 10.1093/infdis/jiaa577.
- [32] A. A. T. M. Bosch *et al.*, “Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study,” *Am. J. Respir. Crit. Care Med.*, vol. 196, no. 12, pp. 1582–1590, Dec. 2017, doi: 10.1164/rccm.201703-0554OC.
- [33] S. M. Teo *et al.*, “The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development,” *Cell Host Microbe*, vol. 17, no. 5, pp. 704–715, May 2015, doi: 10.1016/j.chom.2015.03.008.
- [34] N. H. Vissing, B. L. K. Chawes, and H. Bisgaard, “Increased Risk of Pneumonia and Bronchiolitis after Bacterial Colonization of the Airways as Neonates,” *Am. J. Respir. Crit. Care Med.*, vol. 188, no. 10, pp. 1246–1252, Nov. 2013, doi: 10.1164/rccm.201302-0215OC.
- [35] L. Toivonen *et al.*, “Early nasal microbiota and acute respiratory infections during the first years of life,” *Thorax*, vol. 74, no. 6, pp. 592–599, Jun. 2019, doi: 10.1136/thoraxjnl-2018-212629.
- [36] G. Biesbroek *et al.*, “Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children,” *Am. J. Respir. Crit. Care Med.*, vol. 190, no. 11, pp. 1283–1292, Dec. 2014, doi: 10.1164/rccm.201407-1240OC.

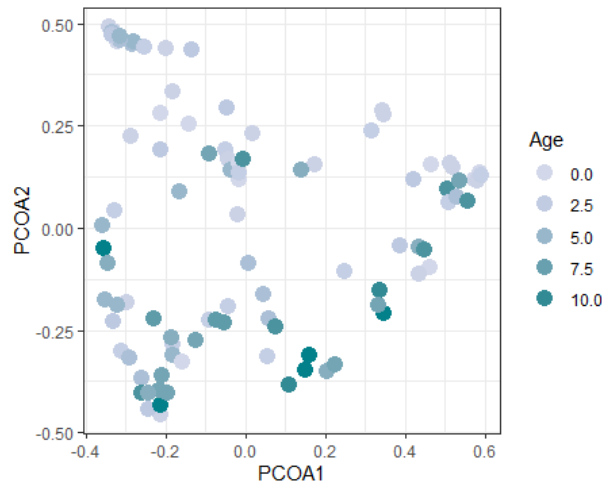
- [37] T. Ma *et al.*, “The diversity and composition of the human gut lactic acid bacteria and bifidobacterial microbiota vary depending on age,” *Appl. Microbiol. Biotechnol.*, vol. 105, no. 21, pp. 8427–8440, Nov. 2021, doi: 10.1007/s00253-021-11625-z.
- [38] K. D. Cephas *et al.*, “Comparative Analysis of Salivary Bacterial Microbiome Diversity in Edentulous Infants and Their Mothers or Primary Care Givers Using Pyrosequencing,” *PLoS ONE*, vol. 6, no. 8, p. e23503, Aug. 2011, doi: 10.1371/journal.pone.0023503.
- [39] M. Dzidic *et al.*, “Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay,” *ISME J.*, vol. 12, no. 9, pp. 2292–2306, Sep. 2018, doi: 10.1038/s41396-018-0204-z.
- [40] C. Hu *et al.*, “Oxidative stress leads to reduction of plasmalogen serving as a novel biomarker for systemic lupus erythematosus,” *Free Radic. Biol. Med.*, vol. 101, pp. 475–481, Dec. 2016, doi: 10.1016/j.freeradbiomed.2016.11.006.

## ANNEXES

### ANNEX I. Respiratory microbiome in RSV



**Figure 1. Patient clustering based on genus-level respiratory microbiome profile at diagnosis. A)** Principal coordinate analysis based on Bray–Curtis dissimilarities showing the nasopharyngeal microbiota composition in RSV patients and controls at diagnosis. Each data point indicates a nasopharyngeal sample coloured by the identified  $k$ -mean cluster (1,2,3,4). Clustering of samples was performed on the community composition dissimilarity (beta diversity) using the  $k$ -means and  $hclust$  R functions. **B).** Stacked bar plot indicating the fraction of RSV and control samples belonging to each of the identified clusters. **C).** Age (months) distribution among the identified clusters. Box plots represent the 25th and 75th percentiles (lower and upper boundaries of boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25th and 75th percentiles; whiskers). Statistical significance was assessed using a Wilcoxon’s test. **D)** Relative abundance of top 10 respiratory genera in the respiratory microbiome for each of the identified clusters. All the remaining respiratory genera with lower abundances were included in the group named “low abundant”.



**Figure 2. Age drives significant changes in the respiratory microbiome at diagnosis of RSV and healthy infants.** Principal coordinate analysis based on Bray–Curtis dissimilarities showing the nasopharyngeal microbiota composition in RSV patients and controls at diagnosis. Each data point indicates a nasopharyngeal sample coloured by age category.  $R^2$  and statistical significance of the association between age and the overall microbiota composition was assessed using PERMANOVA-tests (9999 permutations).

**Table 1. Differentially abundant respiratory genera in RSV compared to healthy infants.** Results of Maaslin analysis to compare CLR-transformed abundances of microbial genera in the respiratory microbiome between RSV and control. Only significant results are shown ( $q$ -value  $<0.05$ ).

Genus	Comparison	FC	q-value
Haemophilus	RSV Diseased vs Control	0.22	0.0000112079
Dolosigranulum	RSV Diseased vs Control	-0.086	0.0101820849
Moraxella	RSV Diseased vs Control	-0.14	0.0371237236
Corynebacterium	RSV Diseased vs Control	-0.093	0.0374414478
Streptococcus	RSV Diseased vs Control	0.075	0.0383803895

**Table 2. Differentially abundant respiratory genera between RSV disease groups compared to healthy infants, by disease severity.** Results of Maaslin analysis to compare CLR-transformed abundances of microbial genera in the respiratory microbiome between RSV and control. Only significant results are shown ( $q$ -value  $<0.05$ ).

Genus	Comparison	FC	q-value
Moraxella	Hospitalised RSV vs Control	-0.24	$6 \times 10^{-5}$
Moraxella	Ventilated RSV vs Control	-0.21	$2 \times 10^{-4}$

**Table 3. Differentially abundant respiratory genera upon RSV recovery.** Results of Maaslin analysis to compare CLR-transformed abundances of microbial genera in the respiratory microbiome between RSV groups before and after recovery (RSV Recovered vs RSV Diseased). Only significant results are shown ( $q$ -value  $<0.05$ ).

Genus	Comparison	FC	q-value
Haemophilus	RSV Diseased vs RSV Recovered	-0.14132621	0.00171439
Moraxella	RSV Diseased vs RSV Recovered	0.08221280	0.04037129
Streptococcus	RSV Diseased vs RSV Recovered	-0.05920760	0.04424302
Corynebacterium	RSV Diseased vs RSV Recovered	0.05643170	0.04429076
Dolosigranulum	RSV Diseased vs RSV Recovered	0.04018092	0.04936118

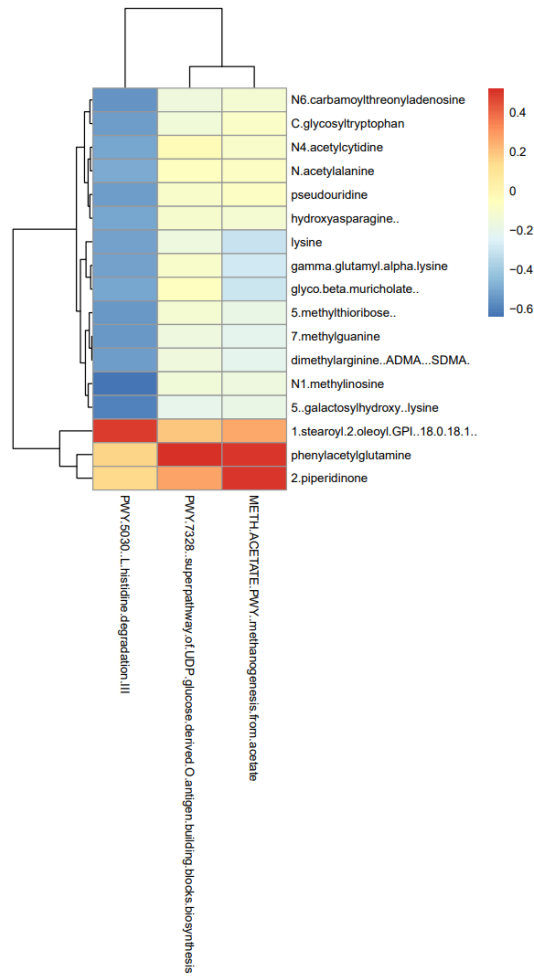
**Table 4. Differentially abundant respiratory genera upon RSV recovery for each RSV group separately.** Results of Maaslin analysis to compare CLR-transformed abundances of microbial genera in the respiratory microbiome of RSV Ventilated group at diagnosis and upon recovery (RSV Recovered vs RSV Diseased). Only significant results are shown ( $q$ -value  $<0.05$ ).

Genus	Comparison	FC	q-value
Haemophilus	Hospitalised RSV Diseased vs RSV Recovered	-0.21	0.0073
Haemophilus	Ventilated RSV Diseased vs RSV Recovered	-0.30	0.0008
Dolosigranulum	Ventilated RSV Diseased vs RSV Recovered	0.14	0.0060
Corynebacterium	Ventilated RSV Diseased vs RSV Recovered	0.15	0.0159

## ANNEX II. Microbiome-metabolite correlations in RSV

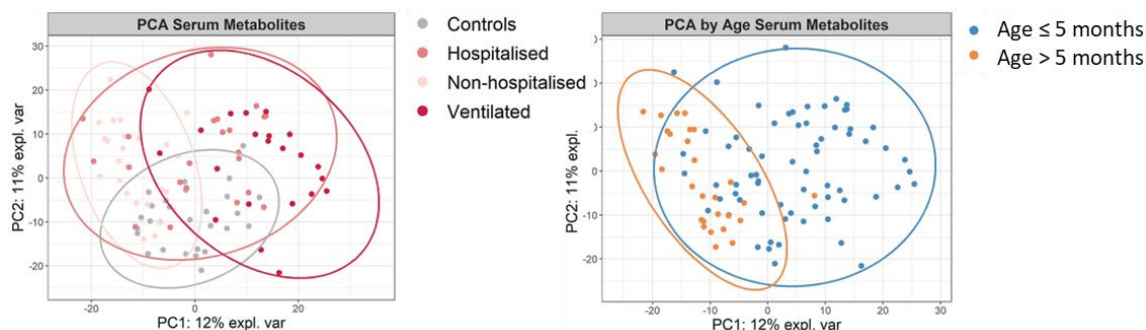
**Table 1. Clusters comparison between respiratory microbiome and serum metabolome for all samples at diagnosis.**

Names	n shared	% shared metabolome	% shared respiratory microbiome
Cluster 1 (metabolome) & Cluster 1 (respiratory microbiome)	5	0.2	0.2
Cluster 2 (metabolome) & Cluster 1 (respiratory microbiome)	12	0.3	0.5
Cluster 3 (metabolome) & Cluster 1 (respiratory microbiome)	6	0.2	0.2
Cluster 1 (metabolome) & Cluster 2 (respiratory microbiome)	4	0.1	0.5
Cluster 2 (metabolome) & Cluster 2 (respiratory microbiome)	3	0.1	0.4
Cluster 3 (metabolome) & Cluster 2 (respiratory microbiome)	1	0.0	0.1
Cluster 1 (metabolome) & Cluster 3 (respiratory microbiome)	4	0.1	0.1
Cluster 2 (metabolome) & Cluster 3 (respiratory microbiome)	17	0.4	0.4
Cluster 3 (metabolome) & Cluster 3 (respiratory microbiome)	15	0.5	0.4
Cluster 1 (metabolome) & Cluster 4 (respiratory microbiome)	13	0.5	0.5
Cluster 2 (metabolome) & Cluster 4 (respiratory microbiome)	4	0.1	0.2
Cluster 3 (metabolome) & Cluster 4 (respiratory microbiome)	6	0.2	0.2



**Figure 1. Correlations between stool microbial pathways and serum metabolome.** Heatmap showing the Spearman's coefficient of the correlation between stool microbial pathways and serum metabolites. Red indicates a positive correlation, and blue a negative correlation. Correlations were performed using Spearman correlation and Benjamini-Hochberg adjustment for multiple comparison. Significant correlations were identified for those with  $p$ -adjusted  $<0.05$ . For visualisation purposes, only stool microbial pathways and serum metabolites with at least one significant correlation with Spearman's with  $p$ -value  $<0.05$  are shown.

## ANNEX III. Serum metabolomics analyses



**Figure 116. PCA on serum metabolites.** PCA with unit variance scaling. A. Dots are coloured by groups: controls in grey and non-hospitalised, hospitalised and ventilated from light to dark pink. B. Dots are coloured by age: infants  $\leq 5$  months in blue and  $>5$  months in orange.

**Table 1.** Top discriminant metabolites (absolute value of loadings of component 1  $>0.06$ ) in the PLS-DA with controls, RSV non-hospitalised, hospitalised and ventilated infants. Loading values in components 2 and 3 are also mentioned. Loadings correspond to the weight of each original variable's contribution to the corresponding latent component.

Metabolite Name	Loading comp 1	Loading comp 2	Loading comp 3
cystein-S-yl-acetaminophen	0.087	0.004	0.037
acetaminophen sulfate	0.087	0.030	0.018
methoxyacetaminophen sulfate	0.086	0.009	0.039
acetamidophenol	0.086	0.009	0.035
vanillic alcohol sulfate	0.083	-0.004	0.014
N acetyl cystein S yl acetaminophen	0.083	0.027	0.050
acetamidophenylglucuronide	0.082	0.021	0.046
N formylanthranilic acid	0.081	0.006	0.047
methylthio acetaminophen sulfate	0.079	0.041	0.017
methoxyacetaminophen glucuronide	0.078	0.017	0.062
hydroxyacetaminophen sulfate	0.077	0.024	0.022
diglycerol	0.076	-0.011	0.009
aminophenol sulfate 2	0.074	0.022	0.023
N1 N12 diacetylspermine	0.074	0.007	0.004
acetylspermidine	0.072	0.005	0.024
25790	0.070	0.001	-0.042
12798	-0.070	0.009	-0.021
stearoyl 2 arachidonoyl GPE 18 0 20 4 sulfate	0.070	0.032	0.008
glucuronate	-0.067	0.014	-0.055
palmitoyl 2 arachidonoyl GPE 16 0 20 4	0.067	-0.041	-0.012
acetylarginine	0.066	0.031	0.001
carnitine	-0.066	0.010	-0.001
ureidopropionate	-0.066	0.009	0.008
stearoyl 2 docosahexaenoyl GPE 18 0 22 6	0.065	-0.020	-0.041
enyl stearoyl 2 linoleoyl GPC P 18 0 18 2	0.065	0.033	-0.013
S methylcysteine sulfoxide	-0.065	0.019	0.025
threonate	-0.064	0.011	0.000
stearoyl 2 arachidonoyl GPI 18 0 20 4	-0.064	-0.017	-0.009
stearoyl 2 dihomo linolenoyl GPE 18 0 20 3n3 or 6	-0.063	0.007	0.035
palmitoyl 2 eicosapentaenoyl GPE 16 0 20 5	0.062	0.018	-0.009
aminobutyrate	0.062	0.020	-0.011
oxalate ethanedioate	-0.062	0.022	0.012
palmitoyl 2 docosahexaenoyl GPE 16 0 22 6	-0.062	0.003	-0.035
	0.060	0.029	-0.013

**Table 2.** Top discriminant metabolites (absolute value of loadings of component 1 >0.06) in the PLS-DA with all three RSV groups (Non-hospitalised, Hospitalised, Ventilated). Loading values of component 2 are also mentioned. Loadings correspond to the weight of each original variable's contribution to the corresponding latent component.

Metabolite	Loading comp 1	Loading comp 2
glucuronate	0.074	-0.013
1-methyladenosine	0.067	0.024
23641	0.067	-0.004
12117	0.067	-0.005
vanillicalcoholsulfate	0.067	0.013
3-cysteinylacetaminophen	0.067	0.041
21410	0.066	0.048
21-hydroypregnenolone monosulfate 2	0.066	0.034
4-acetamidobutanoate	0.065	-0.028
nervonylcarnitine C-241	-0.065	0.022
1-stearoyl2docosapentaenoyl GPCO180225 n3	-0.065	0.016
Androsteroidmonosulfate C19H28O6S1	0.064	0.048
N1-methylinosine	0.064	0.020
pseudouridine	0.063	0.001
salicylate	-0.063	0.003
2-methoxyacetaminophensulfate	0.063	0.045
1-stearoyl2docosaheanoyl GPI180226	-0.062	-0.008
Smethylcysteinesulfoide	-0.062	0.002
1,1-enylstearoyl2docosaheanoyl GPEP180226	-0.062	0.026
trigonellineNmethylnicotinate	-0.062	-0.046
1,1-enylstearoyl2docosaheanoyl GPCP180226	-0.062	0.021
24582	0.062	-0.071
21364	0.062	0.012
tryptophan	-0.061	0.047
4-acetamidophenol	0.061	0.038
diglycerol	0.061	0.006

**Table 3.** Top discriminant metabolites (absolute value of loadings of component 1 >0.06) in the PLS-DA with controls vs RSV-positive infants. Loading values of component 2 are also mentioned. Loadings correspond to the weight of each original variable's contribution to the corresponding latent component.

Metabolite	Loading comp 1	Loading comp 2
4-acetaminophen sulfate	-0.089	0.063
3-methylthio-acetaminophen sulfate	-0.088	0.055
3-N-acetyl-cystein-S-yl-cetaminophen	-0.085	0.054
4-acetamidophenylglucuronide	-0.081	0.048
24576	0.080	-0.086
2-hydroxyacetaminophen-sulfate	-0.078	0.047
4-acetamidophenol	-0.078	0.026
2-methoxyacetaminophen-sulfate	-0.077	0.024
3-cystein-S-yl-acetaminophen	-0.077	0.029
1-stearoyl-2-arachidonoyl-GPE 18 0 20 4	-0.076	0.003
2-methoxyacetaminophen-glucuronide	-0.075	0.042
4-aminophenol-sulfate-2	-0.075	0.052
gamma-glutamylcitrulline	0.074	0.011
24456	-0.073	-0.021
1-palmitoyl-2-arachidonoyl GPE-16-0-20-4	-0.073	-0.012
1-stearoyl-2-docosaheanoyl GPE-18-0-22-6	-0.073	0.013
N-formylanthranilic-acid	-0.072	0.021
24578	0.070	-0.086
1-stearoyl-2-linoleoyl GPE 18-0-18-2	-0.070	0.025
1-palmitoyl-2-linoleoyl GPE 16-0-18-2	-0.069	0.009
24549	0.069	0.018
5-methyluridine-ribothymidine	0.069	-0.054
vanillic-alcohol-sulfate	-0.069	0.019
spermidine	-0.068	-0.006
gamma-glutamylphenylalanine	-0.068	-0.025
15674	0.067	0.037
1-palmitoyl-2-oleoyl-GPE--16-0-18-1	-0.067	-0.014

N1-N12-diacetylspermine	-0.067	0.005
1-methylnicotinamide	0.066	-0.082
1-palmitoyl-2-docosaheanoyl-GPE--16-0-22-6	-0.066	-0.002
7-methylanthine	-0.066	-0.027
methionine-sulfoide	-0.066	-0.032
N-palmitoyl-sphinganine--d18-0-16-0	-0.066	0.008
N-lactoyl-phenylalanine	-0.065	-0.002
1-stearoyl-2-oleoyl-GPE--18-0-18-1	-0.065	0.020
isocitrate	0.065	0.023
fibrinopeptide A	-0.065	0.028
N-acetylmethionine-sulfoide	-0.065	-0.053
N-1-acetylspermidine	-0.065	-0.002
1-palmitoyl-2-eicosapentaenoyl GPE-16-0-20-5	-0.064	-0.015
1-oleoyl-2-linoleoyl GPE-18-1-18-2	-0.064	0.026
N-lactoyl-isoleucine	-0.063	0.005
N-stearoyl-sphingosine d18-1-18-0	-0.063	-0.002
threonate	0.063	-0.002
12127	-0.063	0.062
1-stearoyl-2-dihomo-linolenoyl-GPE 18-0-20 3n3 or 6	-0.062	0.005
Lactose	-0.062	0.005
N-stearoyl-sphinganine d18-0-18-0	-0.062	0.016
N-lactoyl-valine	-0.061	-0.012
pyridoal	0.060	-0.023
N-palmitoyl-sphingosine d18-1-16-0	-0.060	0.007

**Table 3.** Metabolites showing a significant difference between controls, non-hospitalised, hospitalised and ventilated infants (ANOVA between the four groups, ANOVA only with RSV groups and Welch t-test between controls and RSV-positive). Median, interval quantile range (IQR), p-value and q-value are mentioned.



annex%20metabolom  
ics%20univariate%20s

**Table 4.** Top metabolites by abundance for each of the metabolome clusters. Median abundance of the top 20 metabolites in each of the identified metabolome clusters are presented.

Metabolite	Median abundance	Metabolome cluster
methyl 4 hydroxybenzoate sulfate	100.09	Cluster 1
diglycerol	20.37	Cluster 1
saccharin	10.94	Cluster 1
21364	6.32	Cluster 1
24544	3.12	Cluster 1
12096	2.81	Cluster 1
aminoheptanoate	2.64	Cluster 1
hydroxypregnenolone monosulfate 2	2.49	Cluster 1
glucuronate	2.16	Cluster 1
methoxyacetaminophen glucuronide	2.14	Cluster 1
25790	2.12	Cluster 1
maltotriose	2.11	Cluster 1
24571	2.11	Cluster 1
alpha hydroxypregnenolone 3 sulfate	2.08	Cluster 1
indoxyl sulfate	2.07	Cluster 1
23641	1.94	Cluster 1
12112	1.94	Cluster 1
eicosapentaenoylglycerol 20 5	1.91	Cluster 1
acetamidophenol	1.90	Cluster 1
trans urocanate	84.15	Cluster 2
3 amino 2 piperidone	22.82	Cluster 2
pyroglutamylvaline	17.46	Cluster 2
18901	13.02	Cluster 2
lactoyl isoleucine	11.78	Cluster 2
25656	10.30	Cluster 2

lactoyl valine	10.05	Cluster 2
cis urocanate	8.94	Cluster 2
dexpanthenol	5.80	Cluster 2
beta cryptoxanthin	5.78	Cluster 2
triethanolamine	4.96	Cluster 2
N lactoyl phenylalanine	4.93	Cluster 2
pyroglutamylalanine	4.41	Cluster 2
N lactoyl tyrosine	4.16	Cluster 2
hydroxyphenylacetyl carnitine	4.05	Cluster 2
25343	3.82	Cluster 2
oxoproline	3.63	Cluster 2
N lactoyl leucine	3.58	Cluster 2
gamma glutamylhistidine	3.13	Cluster 2
pregnenolone sulfate	2.72	Cluster 3
delta CEHC	2.48	Cluster 3
valylglycine	2.32	Cluster 3
24549	2.26	Cluster 3
methylnicotinamide	2.21	Cluster 3
taurocholate sulfate	2.19	Cluster 3
glycocholate sulfate	2.14	Cluster 3
gamma glutamylcitrulline	2.11	Cluster 3
hyocholate	2.08	Cluster 3
taurochenodeoxycholic acid 3 sulfate	2.05	Cluster 3
15674	2.02	Cluster 3
leucylalanine	2.00	Cluster 3
isoleucylglycine	2.00	Cluster 3
5 dichloro 2 6 dihydroxybenzoic acid	1.98	Cluster 3
gamma CEHC	1.96	Cluster 3
valylleucine	1.80	Cluster 3
glycochenodeoxycholate 3 sulfate	1.79	Cluster 3
12798	1.77	Cluster 3
24556	1.77	Cluster 3