Oropharyngeal microbiota clusters in children with asthma/wheeze associate with allergy, blood transcriptomic immune pathways and exacerbations risk

Mahmoud I. Abdel-Aziz, PhD[1,2,3,4]; Jonathan Thorsen, MD, PhD[5,6]; Simone Hashimoto, MD, PhD[1,2,3,7]; Susanne J. H. Vijverberg, PhD[1,2,3]; Anne H. Neerincx, PhD[1,2,3]; Paul Brinkman, PhD[1,2,3]; Wim van Aalderen, MD, PhD[7]; Jakob Stokholm, MD, PhD[5,8]; Morten Arendt Rasmussen, MD, PhD[5,8]; Michael Roggenbuck-Wedemeyer, MD, PhD[9,10]; Nadja H. Vissing, MD, PhD[5]; Martin Steen Mortensen, PhD[9]; Asker Daniel Brejnrod, PhD[11]; Louise J. Fleming, MD[12]; Clare S. Murray, MD[13]; Stephen J. Fowler, MD, PhD[13]; Urs Frey, MD, PhD[14]; Andrew Bush, MD, FRCP, FRCPCH[12]; Florian Singer, MD, PhD[15]; Gunilla Hedlin, MD, PhD[16]; Björn Nordlund, RN, PhD[16]; Dominick E. Shaw, PhD[17]; Kian Fan Chung, MD, DSc[12]; Ian M. Adcock, PhD[12]; Ratko Djukanovic, MD, PhD[18]; Charles Auffray, PhD[19]; Aruna T. Bansal, PhD[20]; Ana R. Sousa, PhD[21]; Scott S. Wagers, MD, PhD[22]; Bo Lund Chawes, MD, PhD[6]; Klaus Bønnelykke, MD, PhD[6]; Søren Johannes Sørensen, PhD[9]; Aletta D. Kraneveld PhD[23]; Peter J. Sterk, MD, PhD[1,2]; Graham Roberts, DM[18]; Hans Bisgaard, MD, PhD[5]; Anke H. Maitland-van der Zee, PharmD, PhD[1,,2,3,7]; on behalf of the U-BIOPRED Study Group.

1 Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands.

2 Amsterdam Institute for Infection and Immunity, Amsterdam, the Netherlands.

3 Amsterdam Public Health, Amsterdam, the Netherlands.

4 Department of Clinical Pharmacy, Faculty of Pharmacy, Assiut University, 71526 Assiut, Egypt.

5 COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

6 Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

7 Department of Paediatric Pulmonary Medicine, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands.

26 8 Department of Food Science, University of Copenhagen, Frederiksberg, Denmark.
27 9 Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
28 10 Novozymes, Bagsvaerd, Denmark.
29 11 Section of Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby,
30 Denmark.
31 12 National Heart and Lung Institute, Imperial College London, and Royal Brompton and Harefield NHS Trust,
32 London, United Kingdom.
33 13 Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology,
34 Medicine and Health, University of Manchester, and Manchester Academic Health Science Centre and NIHR
35 Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom.
36 14 University Children's Hospital Basel, University of Basel, Basel, Switzerland.
37 15 Division of Paediatric Pulmonology and Allergology, Department of Paediatrics and Adolescent Medicine,
38 Medical University of Graz, Austria & Division of Paediatric Respiratory Medicine and Allergology, Department of
39 Paediatrics, Inselspital, Bern University Hospital, University of Bern, Switzerland.
40 16 Astrid Lindgren Children's Hospital, Karolinska University Hospital, and Department of Women's and Children's
41 Health, Karolinska Institutet, Stockholm, Sweden.
42 17 NIHR Respiratory Biomedical Research Unit, University of Nottingham, Nottingham, United Kingdom.
43 18 NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust and
44 Clinical and Experimental Sciences and Human Development and Health, University of Southampton, Southampton,
45 United Kingdom.
46 19 European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon, France.
47 20 Acclarogen Ltd, St John’s Innovation Centre, Cambridge, United Kingdom.
48 21 Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park, United Kingdom.
49 22 BioSci Consulting, Maasmechelen, Belgium.
50 23 Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University,
51 Utrecht, the Netherlands.
52

Corresponding author details:

Prof. Dr. Anke-Hilse Maitland-van der Zee

Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands.

e-mail: a.h.maitland@amsterdamumc.nl, Telephone: +31 (0) 20 5668137

Authorship contribution:

MIA has performed the analysis and drafted the initial version of the manuscript. MIA, AHM, JT, GR and PJS have contributed to the design of the analysis plan. All co-authors have contributed to the acquisition of data, interpretation of the analysis, revision, drafting, critical appraisal and ensuring accuracy and integrity of the analysis. All co-authors have provided final approval of the version to be published. AHM is the corresponding author.

Funding:

U-BIOPRED has received funding from the Innovative Medicines Initiative (IMI) Joint Undertaking under grant agreement no. 115010, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contributions (www.imi.europa.eu). The researcher MIA has received a full scholarship from the Ministry of Higher Education of the Arab Republic of Egypt.

Short running head: "Bacterial asthma/wheezing subtypes in children"

Manuscript Subject descriptor: PEDIATRICS 14.1 Clinical Studies: Asthma

Article type: Original article

Word count: 4,000 without abstract, references, tables, figures, and disclosures

References number: 60

Tables: 4 in the main manuscript + 7 supplementary

Figures: 3 in the main manuscript + 17 supplementary

Online Repository: This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org.

At a Glance Commentary

Scientific Knowledge on the Subject

Children with asthma or wheezing were reported to have airway microbial imbalances. However, it is not yet clear whether characterizing the microbiota in oropharyngeal swabs, as a non-invasive sampling compartment, could help identifying clinically relevant phenotypes in children with asthma or wheezing.

What This Study Adds to the Field

In this study, we show that the oropharyngeal microbiota could identify 4 distinct clusters that were different in allergic status and spirometry and that exhibited differential expression of TGF-β and Wnt/β-catenin signaling pathways in blood. These clusters were independent predictors of future exacerbation events in children with severe disease, within 12-18 months of follow-up.

Abstract

Rationale:

Children with preschool wheezing or school-age asthma are reported to have airway microbial imbalances.

Objective:

To identify clusters in children with asthma or wheezing using oropharyngeal microbiota profiles.

Methods:

Oropharyngeal swabs, from the Unbiased BIOmarkers for the Prediction of REspiratory Disease outcomes pediatric asthma/wheezing cohort, were characterized by 16S rRNA gene sequencing and unsupervised hierarchical clustering was performed on the Bray-Curtis β-diversity. Enrichment scores (ESs) of the MSigDB Hallmark gene sets were computed from the blood transcriptome using gene set variation analysis. Children with severe asthma or severe wheezing were followed up for 12-18 months, with assessing the frequency of exacerbations.

Measurements and Main Results:

Oropharyngeal samples of 241 children (age range: 1-17 years, 40% female) revealed 4 taxa-driven clusters dominated by *Streptococcus*, *Veillonella*, *Rothia* and *Haemophilus*, respectively. The clusters showed significant differences in atopic dermatitis, grass pollen sensitization, FEV_1 % predicted post-salbutamol, and the annual asthma exacerbation frequency during follow-up. The *Veillonella*-cluster was the most allergic and included highest percentage of children with ≥2 exacerbations/year during follow-up. The oropharyngeal clusters were different in the ESs of transforming growth factor β (highest in *Veillonella*-cluster) and Wnt/β-Catenin signaling (highest in *Haemophilus*-cluster) transcriptomic pathways in blood (all *q*-values < 0.05).

Conclusion:

The analysis of the oropharyngeal microbiota of children with asthma/wheezing identified four clusters with distinct clinical characteristics (phenotypes) that associate with exacerbations' risk and transcriptomic pathways involved in airway remodeling. This suggests that further exploration of the oropharyngeal microbiota may lead to novel pathophysiological insights and potentially new treatment approaches.

Abstract word count: 250

Keywords (Mesh): Asthma; Wheezing; Phenotype; Microbiota; Precision Medicine

129 **Introduction**

130 Asthma is the most common chronic disease affecting children. Childhood asthma is a multi-faceted
131 disease that presents with varied clinical manifestations and treatment responses. In addition, childhood
132 asthma-like symptoms were reported to increase risk of COPD in adulthood [1]. The underlying
133 mechanisms of childhood asthma or preschool wheezing are not completely understood and thought to
134 originate from multiple interacting factors, including genetic susceptibility, environmental exposures,
135 and microbiota [2]. The host-residing microbiota regulates and works in an alliance with the immune
136 system and is thought to partly mediate the link between environmental exposures and asthma
137 pathophysiology [3, 4]. Therefore, assessing the microbiota profiles in children with wheezing or asthma
138 may reveal distinct host, microbiome, and disease links that could play a role in disease development.

139 Omics-guided asthma characterization approaches have been utilized recently to delineate groups (i.e.
140 clusters) of patients with asthma that share similar molecular profiles and exhibit distinct phenotypic
141 asthma characteristics [5]. Underpinning these asthma phenotypes with specific molecular fingerprints
142 may lead to tailoring diagnostic and/or therapeutic decisions. This may result in better patient care,
143 reduce the risk of severe complications, and prevent inappropriate medication use. In adults, the
144 Unbiased BIOmarkers for the Prediction of REspiratory Disease outcomes (U-BIOPRED) revealed sputum
145 taxa-driven clusters linked to asthma severity and underlying inflammatory status, particularly sputum
146 neutrophilia [6]. Similar findings in adults with asthma have been reported elsewhere [7], suggesting that
147 the sputum microbiota may help refine the neutrophilic asthma phenotype in adult patients with severe
148 asthma. Combining other omics, including transcriptomics, delineates associated molecular mechanisms
149 characterizing the taxa-driven clusters and may improve our understanding of the pathophysiology [8].
150 In children, obtaining samples representing the lower airways (e.g. induced sputum, bronchoalveolar
151 lavage fluid, or endobronchial biopsies samples) is challenging. Evidence suggests the ecological

152 continuity of the microbiota structure within the respiratory tract [9-11], advocating the need to
153 investigate the value of non-invasive microbiota sampling in the assessment of children with wheeze or
154 asthma. Studies have shown changes in pharyngeal microbiota are associated with the risk of later
155 developing asthma in young children [12-15] or associated with different asthma characteristics [16, 17].
156 However, attempts to assess the value of airways microbiota in stratifying children with wheezing or
157 asthma into clinically meaningful groups are scarce.

158 To clarify the role of the oropharyngeal microbiota in pediatric preschool-age wheeze and school-age
159 asthma, we examined oropharyngeal swabs by 16S rRNA gene sequencing in the European U-BIOPRED
160 pediatric cohort [18]. In a previous report from this dataset, we found limited differences in diversity and
161 differential abundances between children with severe and mild-to-moderate asthma or wheeze, but a
162 significant association with age [19]. Building on these results, we hypothesized that oropharyngeal
163 microbiota profiles may reveal distinct clusters in children with preschool-age wheezing or asthma
164 during school-age. This study specifically aims to 1) perform unsupervised unbiased clustering of
165 oropharyngeal microbiota profiles of children and assess whether these clusters are linked with district
166 asthma characteristics (phenotypes), 2) elucidate the possible underlying biological pathways of the
167 revealed phenotypic clusters using the peripheral blood transcriptome, and 3) to reveal whether the
168 baseline clusters are associated with future exacerbations during 12-18 months of follow-up in children
169 with severe school-age asthma and severe preschool-age wheezing.

170

171 **Methods**

172 **Study design**

173 The U-BIOPRED study is a multicenter European observational cohort study including children with
174 physician-diagnosed mild-to-moderate and severe school-age asthma and preschool wheezing aged 1-17
175 years old as previously described [18]. Children were recruited from 7 study centers encompassing 5
176 European countries (United Kingdom, Denmark, Sweden, the Netherlands, and Switzerland) and were
177 mainly of a Caucasian ethnic origin. Children were allocated into four cohorts comprising A) severe
178 school-age asthma, B) mild-to-moderate school-age asthma, C) severe preschool-age wheeze or D) mild-
179 to-moderate preschool-age wheeze . All study centers obtained ethics committees approval and
180 parents/caregivers gave written informed consent. Children gave assent where age-appropriate. The
181 study was registered at ClinicalTrials.gov (identifier number: NCT01982162). All study centers followed
182 uniform Standard Operating Procedures (SOPs) in all study procedures as described [18].

183 **Participants**

184 A total of 131 preschool-aged children with wheeze (n=77 severe, n=54 mild-moderate) and 140 school-
185 aged children with asthma (n=97 severe, n=43 mild-moderate) were recruited to the U-BIOPRED
186 pediatric cohorts. The flow diagram of patient inclusion is shown in Figure E1. Asthma/wheeze severity
187 was defined according to the Global Initiative for Asthma (GINA) and the Innovative Medicines Initiative
188 guidelines [20, 21] and fully described elsewhere [18]. Body mass index (BMI) was converted into z-
189 scores according to the World Health Organization growth charts [22]. Participants were assessed for
190 atopy [23, 24] (details in the online supplement) and environmental tobacco smoke exposure by
191 measuring urinary cotinine levels [18]. In addition, they underwent spirometry and fractional exhaled
192 nitric oxide (FE_{NO}) testing (for older capable children) as described [18]. Quality of life was evaluated
193 using the standard Pediatric Asthma (Caregiver's) Quality of Life Questionnaire (PA(C)QLQ) [25, 26] and

194 asthma control was evaluated using the (childhood) Asthma Control Test ((c)ACT) [27, 28]. A (c)ACT score
195 >19 indicates well-controlled asthma. Other family and medical histories were collected from the
196 children or caregivers and the physicians as appropriate.

197 Children with severe asthma or severe wheezing (n= 125/151, 82.8%) were followed up prospectively in
198 time (12-18 months after the baseline visit), during which the frequency of asthma exacerbation (as
199 defined [29]) per year was documented (i.e. future exacerbations).

200 **Samples collection and omics analysis**

201 Full details for samples collection and omics analysis are described previously [19] and in the online
202 supplement.

203 **Oropharyngeal swabs and 16S gene V4 microbiome sequencing**

204 Briefly, oropharyngeal (throat) swabs for microbiota processing were taken from the posterior
205 pharyngeal wall, behind the uvula. The 16S ribosomal RNA (rRNA) gene in the collected samples was
206 amplified for the hypervariable V4 region and sequencing was performed using the Illumina MiSeq
207 Desktop Sequencer. The microbiome sequencing data can be found in the European Nucleotide Archive
208 database (accession number: PRJEB47973). Paired-end reads (250 bp x 2) from amplicon sequencing
209 were quality checked using FastQC [30] and MultiQC [31], and then cleaned by removing V4 region
210 primers using Cutadapt [32]. Further processing was performed using the DADA2 pipeline as described
211 [33]. Briefly, the pipeline works by performing quality filtering and trimming, dereplicating sequences,
212 learning dataset-specific error rates, denoising by removing potentially containing errors sequences,
213 merging paired-end reads while removing mismatches to reduce errors, constructing amplicon sequence
214 variants (ASVs), removing chimera by implementing “bimera” method. Finally, taxonomic classification of
215 ASVs was done by the Silva database version 132 [34].

216 **Blood collection and transcriptomics analysis**

217 Peripheral blood was isolated and gene expression was assessed by Affymetrix HT HG-U133+ PM arrays
218 (Affymetrix, Santa Clara, CA). The microarray data are deposited in the Gene Expression Omnibus
219 database (accession number: GSE123750). Raw microarray data were quality assessed and pre-
220 processed by robust multi-array average normalization (RMA) method [35] as implemented in the affy R
221 package. Transcriptomic pathway enrichment analysis was performed by an unsupervised approach
222 using the gene set variation analysis (GSVA) R package against gene sets from the Molecular Signatures
223 Database (MSigDB) release version 7.4 [36].

224 **Data and statistical analysis**

225 The general data analysis workflow is shown in Figure E2 and is further described previously [6] and in
226 detail in the online supplement.

227 Briefly, a clustering benchmarking strategy using hierarchical ward2, partition around the medoids, and
228 topological data analysis (TDA) was implemented to subtype children with asthma or wheezing according
229 to their microbiota profiles. The clusters were checked for differences in microbial profiles by Shannon α-
230 diversity (using a Kruskal-Wallis H test followed by a Dunn's post-hoc test) and by Linear discriminant
231 analysis Effect Size (LEfSe) for differentially abundant bacteria at phylum and species levels (using an LDA
232 score>2 and an FDR q-value<0.01). Furthermore, the taxa-driven clusters were tested for differences in
233 certain demographic and clinical characteristic (such as allergic status, atopy, and spirometry) and the
234 frequency of future exacerbations/year using Kruskal-Wallis H or Pearson Chi-Square tests as
235 appropriate. Clinical characteristics that were statistically significant after univariate analysis, were
236 adjusted for several potential confounders using logistic or linear multiple regression models as
237 appropriate. Potential confounders (covariates) to adjust for in the regression models were determined
238 using a directed acyclic graph [37] (DAG, Figure E3). A simple mediation analysis was conducted to

239 investigate whether microbiota Cluster 2 is an independent predictor of future asthma exacerbations
240 (≥2) or if this is mediated by the atopic dermatitis status.

241 The GSVA enrichment scores (ES) of the tested gene sets (i.e. transcriptomic pathways) were compared
242 between the clusters by Kruskal-Wallis H-test and multiple testing was corrected by FDR, followed by
243 permutation analysis (10000 times). Linear regression models were used to investigate whether the
244 enrichment scores of the significant gene sets are dependent on the participants' clusters assignments
245 after adjusting for multiple covariates (as defined in the DAG).

246 All analyses were performed using R studio (version 1.1.453) with R software (version 3.6.1).

247

248 **Results**

249 ***Baseline characteristics of study participants***

250 A total of 241 participants provided baseline oropharyngeal swabs that have been sequenced to a read
251 depth of at least 1000 that passed the quality control. The baseline characteristics of the subjected are
252 shown in Table 1. The median age of the included participants was 6 years, 40% were females, 80% were
253 White and 70% were atopic.

254 **Four main taxa-driven clusters were generated**

255 The median sequencing reads number passing the quality control was 11,975 (IQR= 7,613-21,098) per
256 sample. Negative extraction controls showed different microbial composition and significantly lower
257 Shannon α-diversity compared with oropharyngeal swabs, while correct classification of all bacterial
258 genera was achieved in the mock community (Figure E4). The main bacterial genera detected in the
259 oropharyngeal swab samples were; *Streptococcus*, *Veillonella*, *Haemophilus*, *Prevotella*, and *Rothia*
260 (Figure E5).

261 Clustering performed on Bray-Curtis beta-diversity suggested 4 optimum clusters as evaluated by the
262 majority vote of multiple indices (Figure E6), resulting in 4 children groups by hierarchical ward2 (Figure
263 1A) and partition around medoids (PAM, Figure 1B) clustering. Quantitative assessment of similarity in
264 participants' assignment between the hierarchical ward and PAM clustering was performed by means of
265 Rand index (RI= 0.85) and Pearson Chi-Square test (χ^2= 498.1, $p < 1 \times 10^{-4}$) suggesting high similarity in
266 the two clustering assignments. Visual representation by topological data analysis (TDA) shows that the 4
267 microbiota clusters are enriched (according to SAFE scores) into 4 distinct regions of the TDA network
268 (Figure E7A and Figure E7B). The 4 clusters were also visualized by the PCoA plot on the Bray-Curtis
269 distance with depicting the 95% ellipses (Figure E8).

270 The hierarchical clusters had significantly different microbial Shannon α-diversity (Figure E9), showing a
271 decreasing order of median diversity as follows: Cluster 4>Cluster 2>Cluster3=Cluster1. The mean
272 relative abundances of most abundant bacterial genera and the individual samples' compositions were
273 stratified according to the clusters' assignments (Figure E10 and Figure E11, respectively). Analysis by the
274 LEfSe method revealed the differential abundance of the most abundant bacterial phyla and species
275 between the clusters (Figure 2). Cluster 1 showed main enrichment of bacterial genus *Streptococcus* and
276 phylum *Firmicutes*, Cluster 2 of genus *Veillonella* and phylum *Patescibacteria*, Cluster 3 of genus *Rothia*
277 and phylum *Actinobacteria*, and Cluster 4 of genera *Haemophilus* and *Neisseria, as well as phylum*
278 *Proteobacteria*. The TDA was consistent with the LEfSe method, in which the same regions of the TDA
279 network enriched with the cluster's assignment, showed increased relative abundances of the above-
280 mentioned bacterial genera (Figure E12 and Figure E13). Of note, *Moraxella* was detected in a small
281 percentage of samples and in low abundances (5.0% of samples, 0.01% mean relative abundance). It was
282 relatively higher in abundance in Cluster 1 compared with the other clusters, although the results were
283 not statistically significant after FDR correction. Decision tree analysis showed that the relative
284 abundances of only 3 bacterial genera (*Streptococcus*, *Veillonella,* and *Haemophilus*) (Figure E14) could
285 be used to classify the children into the corresponding clusters with an accuracy of 0.86 (95% CI: 0.75-
286 0.94, *P*-value = 8.24×10^{-15}) using a holdout validation set.

287 There were no statistically significant associations between the participants' clusters assignments and
288 the 7 study centers from which samples were collected or the season of sample collection (p-values were
289 0.390 and 0.526, respectively).

290

291 **Taxa-driven clusters show distinct clinical characteristics**

292 Table 2 shows the demographic and clinical characteristics of the 4 taxa-driven clusters. The clusters
293 showed statistically significant differences in atopic dermatitis diagnosis, and FEV_1 % predicted values
294 post-salbutamol (the latter also shown in Figure E15). These differences remained significant after
295 adjusting for several covariates (Table E2, and Table E3, *q*-values <0.05). Table 3 shows characteristics of
296 the atopic sensitization of the clusters. The clusters showed statistically significant differences in
297 (aeroallergen) atopic sensitization specifically to grass pollens and the number of aeroallergen
298 sensitizations (the latter also shown in Figure E16). After adjusting for covariates, grass aeroallergen
299 sensitization (but not overall sensitization) was statistically significantly associated with the clusters
300 (Table E4).

301 **Differential peripheral blood transcriptomic pathways in the taxa-driven clusters**

302 A total of 188/241 (78%) participants provided peripheral blood samples for transcriptomic analysis.
303 Figure 3 shows the statistically significant transcriptomic pathways between the taxa-driven clusters. A
304 total of 3 hallmark pathways were statistically different between the taxa-driven clusters, of which the
305 Wnt/β-Catenin and the Transforming growth factor β (TGF-β) signaling transcriptomic pathways
306 remained significant after multiple testing corrections. Cluster 4 had the highest median ESs for the
307 Wnt/β-Catenin signaling pathway, while cluster 2 had the highest median ESs for the TGF-β signaling
308 pathway. The permutation analysis p-values were 9.9×10^{-5} and 0.001 for the Wnt/β-Catenin and the
309 TGF-β signaling, respectively. The pathways showed statistically significant associations with the clusters
310 after adjusting for different covariates (Table E5).

311

312 **Baseline microbiota clusters were associated with future exacerbations in children with severe asthma**
313 **or wheezing**

314 The baseline microbiota clusters were significantly associated with the exacerbations risk during a
315 follow-up period of 12-18 months (Table 4). Logistic regression analysis showed that microbiota clusters
316 are a significant predictor of having ≥2 future exacerbations/year after adjusting for several covariates
317 (Table E5). In particular, Cluster 3 And Cluster 4 had statistically significantly lower odds ratio (OR) of
318 having ≥2 future exacerbations/year relative to cluster 2, (OR = 0.06, 95% CI: 0.01-0.43, and OR = 0.11,
319 95% CI: 0.03-0.34, *q*-values < 0.05, respectively, Table E6). Other covariates that were associated with
320 frequency of ≥2 exacerbations/year including atopic dermatitis diagnosis, center of inclusion and recent
321 prescription of antibiotics at baseline visit (all *q*-values<0.05, Table E6). Notably, the transcriptomic
322 (Wnt/β-Catenin, and TGF-β signaling) pathways ESs were not selected by the stepwise regression model
323 as significant predictors of having ≥2 future exacerbations. Mediation analysis (Table E7) showed that
324 microbiota Cluster 2 is an independent predictor of having ≥2 future exacerbations and this association
325 was not mediated by the atopic dermatitis status, whether this was not adjusted or adjusted for the
326 different covariates.

327 A summary overview of clinical features associated with the taxa-driven clusters is described in Box 1
328 and Figure E17.

329 **Discussion**

330 Using unsupervised clustering of oropharyngeal microbiota profiles, we found that children with school-
331 age asthma and preschool-age wheezing could be stratified into 4 clusters which differed significantly by
332 1) atopic dermatitis diagnosis, grass pollen sensitization, and post-bronchodilator FEV_1 % predicted, 2)
333 the Wnt/β-Catenin and TGF-β signaling peripheral blood transcriptomic pathways and 3) frequency of
334 future exacerbations among children with severe disease.

335 Age has been reported a significant contributor to the variability in the microbiome compositions, while
336 the largest influence of age on microbial profiles was observed very early in life during which microbial
337 maturation is achieved [38, 39]. In this study, the clustering strategy performed only showed a trend of
338 non-statistically significant differences between the clusters concerning age, with cluster 2 being more
339 likely to include older school-age children compared to the other clusters. In addition, in this study, no
340 clear separation of the oropharyngeal microbiota clusters between children with school-age asthma and
341 children with preschool wheezing stratified by disease severity was found, showing an overlapping
342 oropharyngeal microbiome composition in this studied cohort (aged 1-17 years old). These findings are
343 in line with previous U-BIOPRED study [19], suggesting that classical clinical labelling of asthma or
344 wheezing severity cannot be fully explained by the underlying microbiota composition, and attempts to
345 reclassify/rephenotype children with asthma or wheezing should be sought. In this regard, clustering
346 children by underling microbiota composition may offer new insight in disease pathophysiology,
347 particularly if this could be linked with objective measures of disease burden, such as future
348 exacerbations.

349 The taxa-driven clusters were significantly different in atopic dermatitis diagnosis and grass pollen
350 sensitization. Particularly, Cluster 2 included the highest percentage of children with atopic dermatitis,
351 grass pollen sensitization, and school-aged children with severe asthma. In addition, severe

352 asthma/wheezing children within this cluster had the highest percentage of those who experienced ≥2
353 exacerbations in the following year. Atopic sensitization and co-morbid allergy are amongst the main
354 characteristics that shape childhood asthma phenotypes [40, 41]. Our findings are partly in line with
355 other studies, for example; in one study it was found that concomitant atopic dermatitis might be one of
356 the factors associated with severe asthma [42]. Moreover, reports have shown that atopic sensitization
357 to inhalant allergens increases the risk of acute exacerbations and asthma-related hospital admissions in
358 childhood asthma [43]. In a study conducted in children with asthma aged 3-17 years old, it was found
359 that viral infections and allergen sensitization synergistically increased risk of asthma-related hospital
360 admissions [44]. In our study, both the taxa-driven clusters additionally with atopic dermatitis diagnosis
361 were found to be significantly associated with the frequency of future exacerbations confirming previous
362 observations. This suggest that microbiota-allergen interaction may play a role in the pathogenesis of
363 childhood asthma [45] and this interaction needs further investigations in future childhood asthma
364 studies.

365 Cluster 2 (*Veillonella*-dominant and the second-highest cluster in Shannon α-diversity) and Cluster 4
366 (*Haemophilus*- and *Neisseria*-dominant and the highest cluster in Shannon α-diversity) included the
367 highest percentage of children with severe asthma/wheezing who experienced episodes of future
368 exacerbations (≥2 and 1 exacerbations, respectively) compared with the other clusters. In addition,
369 Cluster 2 children had lower FEV_1% predicted post-salbutamol compared with the other clusters. These
370 results are partly in line with findings from another study which showed that increased bacterial
371 Shannon α-diversity and increased *Veillonella* and *Neisseria* abundances in hypopharyngeal aspirates
372 were associated with increased duration of asthma-like episodes in preschoolers [46]. In the same study,
373 an augmented effect of azithromycin in decreasing the asthma-like episodes was particularly seen in
374 children with elevated bacterial relative abundances of *Veillonella* and *Neisseria* genera [46]. In other
375 microbiome studies in which the oropharynx was sampled, preschool wheezers had partly increased

376 levels of *Neisseria* genus compared with non-wheezers [47], while children with school-age asthma (6-12
377 years old) had increased abundances of *Haemophilus* genus compared with healthy controls [48]. These
378 findings suggest that the pharyngeal relative compositions of the above-mentioned bacterial genera may
379 play a role in childhood asthma pathogenesis and may be useful in the phenotypic assessment of
380 children with asthma/wheezing.

381 Interestingly, Cluster 3 (*Rothia*-dominant) included the highest percentage of children who did not
382 experience any future exacerbations during the follow-up period. This is in line with findings from the U-
383 BIOPRED adult cohort, in which high abundance *Rothia mucilaginosa* in sputum was associated with a
384 less severe asthma phenotype [6]. Moreover, *Rothia mucilaginosa* was found to be negatively correlated
385 with matrix metalloproteinases (MMP-1, MMP-8, and MMP-9) and pro-inflammatory markers (IL-8, IL-
386 1β) in sputum of adults with bronchiectasis [49]. In vivo/in vitro mechanistic investigation in the same
387 study revealed that *Rothia mucilaginosa* potentially mitigates inflammation via inhibiting the NF-κB
388 pathway activation, which may influence the severity and development of chronic airway diseases [49].
389 These findings suggest that *Rothia* species may have protective effect in adults and children with severe
390 asthma or wheezing.

391 The peripheral blood transcriptome showed statistically significant differences between the clusters in
392 two main pathways: the TGF-β and the Wnt/β-Catenin signaling pathways. The TGF-β pathway was
393 relatively up-regulated in Cluster 2 (the most atopic) and, to a lesser extent, Cluster 4 compared with
394 Cluster 1. The TGF-β signaling pathway has been reported to be involved in airway remodeling in atopic
395 asthma following exposure to an allergen [50] and may play a role in asthma severity [51]. The TGF-β
396 pathway has been reported to regulate the immune system in response to bacterial infections [52]. On
397 the other hand, the Wnt/β-Catenin signaling pathway was relatively up-regulated in the
398 *Haemophilus*/*Neisseria* predominant-Cluster 4 compared with the other clusters. Similarly, the Wnt/β-

399 catenin pathway has been reported to regulate remodeling in asthmatic airways and may play a role in
400 asthma pathogenesis [53-55]. In addition, Wnt signaling is believed to be one of the key pathways
401 involved in host-bacterial pathogen interactions [56]. The Canonical Wnt/β-Catenin signaling pathway
402 was found to be activated after infection with virulent strains of *Haemophilus* species leading to
403 disruption of the epithelial barrier in pigs [57]. In addition, it was recently found that WNT/β-catenin
404 signaling is a key regulator of macrophage phagocytosis after exposing COPD airway epithelium to non-
405 typeable *Haemophilus influenzae* [58]. In light of the findings of this study and evidence of involvement
406 of these pathways in airway remodeling and pathogenesis, further research is needed to explore the
407 interaction of the oropharyngeal microbiota with these pathways in childhood asthma.

408 In this study, we showed that clustering on the oropharyngeal microbiota revealed subtypes of
409 asthma/wheezing children with distinct clinical characteristics (phenotypes). A decision tree based on
410 the relative abundances of only 3 main bacterial genera showed that we can classify the children with an
411 86% accuracy as estimated from a hold-out validation set. These bacterial genera were among the most
412 abundant and prevalent bacterial taxa found in other pediatric cohorts in which the pharynx was
413 sampled for microbiota analysis [46-48]. This suggests that the taxa-driven clustering strategy in this
414 study may be generalizable to other asthma/wheezing cohorts and settings. Phenotyping in children with
415 asthma/wheezing remains challenging particularly due to the difficulty of obtaining invasive samples
416 representing the lower airways and their inflammatory status. The oropharyngeal (throat) swabs are
417 relatively easy to obtain and 16S sequencing is likely to be an affordable option to sequence the bacterial
418 profiles within these samples. Hence, using the oropharyngeal taxa-driven clustering may help clinicians
419 to non-invasively define subgroups among children with asthma/wheezing who may benefit from
420 personalized monitoring and/or treatment options. Yet, validation of findings in external cohorts and
421 different populations remains necessary before findings can be rendered translatable to clinical practice,

422 and to assess whether the airway microbiome can be therapeutically targeted in children with asthma or
423 wheezing [59].

424 Our study has many strengths. First, the adopted analysis strategy is unsupervised and is not driven by a
425 priori hypothesis. Second, the prospective international multi-center design of this study, the relatively
426 large sample size, and the wide age range of the included cohort make the results likely more
427 generalizable and valid than previously reported single-center studies with lower sample sizes. Third, the
428 oropharyngeal samples are easy to collect and convenient for patients and healthcare professionals,
429 making the phenotypic assessment based on oropharyngeal microbiota feasible. Finally, internally
430 validating the findings by using different clustering algorithms makes the statistical analysis more robust.

431 However, there are also limitations. First, the 16S amplicon sequencing technique used is limited in
432 identifying bacterial taxa up to species level and this imposes some hindrance in identifying the
433 pathogenicity of the differentially abundant bacterial genera between the clusters. More sophisticated
434 techniques, such as shotgun metagenomics that enable species or strain level identification of bacteria,
435 may therefore provide a better insight in this regard. However, in another U-BIOPRED study conducted
436 in adult patients with severe asthma, clustering based on 16S amplicon sequencing revealed similar
437 patient clusters to shotgun metagenomics, and this suggests that the affordable 16S amplicon
438 sequencing is potentially applicable in clinical settings where metagenomics cannot be performed [6].
439 Second, we sampled only one compartment for bacterial sequencing (oropharyngeal swabs). Whether
440 other sampling compartments and detecting other microorganisms (e.g. viruses/fungi) will provide
441 additional information needs to be determined. Third, although patients were followed up clinically for
442 12 to 18 months, no additional biological samples were collected to assess the longitudinal shifts in
443 microbial profiles with time. Although the taxa-driven clusters were found to be relatively stable after
444 12-18 months in adults with severe asthma [6], further longitudinal investigations are still needed in

445 children with asthma. Forth, we did not collect some parameters in the U-BIOPRED pediatric cohort, such
446 as birth weight and antenatal (in utero) and perinatal antibiotic treatment history, therefore; we were
447 not able to investigate whether they are associated with oropharyngeal microbiota profiles. Further
448 research is needed to investigate whether these characteristics could influence the oropharyngeal
449 microbiota composition. Fifth, assessing the peripheral blood gene expression may not reflect the
450 airway-centric asthma pathobiology, and therefore; further research is needed to investigate whether
451 the microbiota composition is directly linked to pathophysiologic biological pathways underlying the
452 diseases airways. Sixth, corticosteroids have been reported to influence the composition of the airway
453 microbiome [60]. Although we have adjusted for corticosteroids intake in all regression analysis
454 performed, this may not reflect the systemic levels of corticosteroids and this could influence the
455 oropharyngeal microbiota composition. Finally, using an external cohort for validation is still needed to
456 confirm the present findings.

457 In conclusion, we have shown that oropharyngeal taxa-driven clustering can be used as an unsupervised
458 method of detecting subtypes in children with preschool wheezing and school-aged asthma. These
459 subtypes exhibited differences in allergic status, lung function, blood transcriptomic pathways involved
460 in airway remodeling, and frequency of future exacerbations. These findings suggest that the
461 oropharyngeal microbiota can be used as a non-invasive approach to phenotype children with asthma or
462 wheezing.

463

464 **Acknowledgements**

465 The authors would like to thank all the participating children and their families. In addition, the U-
466 BIOPRED is a consortium effort and we wish to acknowledge the help and expertise of the individuals
467 and groups whose names are mentioned in the U-BIOPRED study group list available in the online
468 repository.

469 **Competing interests:**

470 **MIA, SH, SJHV, PB, JS, MAR, MRW, NHV, MSM, ADB, SJF, UF, AB, GH, DS, CA**, **ATB**, **ARS, BLC,** and **ADK**
471 have no conflicts of interest to disclose. **JT** reports receiving Honoria from AstraZeneca. **AHN** reports
472 receiving a grant from ERANET Systems Medicine and ZonMW [project number: 9003035001]. **WVA** has
473 received honoraria for participating in a medical advisory board for Boehringer Ingelheim BV and for
474 giving lectures for OMRON Healthcare Europe BV. G's University received funding from EU IMI for this
475 research. **LF** has received honoraria and Consulting fees from Astra Zeneca, Sanofi, Adherium, Novartis,
476 Teva and Respiri, outside the submitted work. **CSM** reports grants from National Institute for Health
477 Research, Asthma UK, Innovate UK, Moulton Charitable Foundation, North West lung Centre Charity
478 outside the submitted work; honoraria for lecture fees from GSK and Novartis and participation in
479 advisory boards for Sanofi and Adherium Ltd. **FS** reports financial support from Novartis Pharma
480 Switzerland, Vertex Pharmaceuticals Switzerland, and non-financial support from Chiesi Pharmaceuticals
481 Austria outside the submitted work. **BN** owns stocks in MediTuner AB, a company supporting remote
482 digital management of asthma. **KFC**'s institution received MRC, EPSRC and GSK grants for asthma
483 research, and he reports receiving Honoria from Novartis & AZ and participating in advisory board
484 meetings for Roche, Merck, Rickett- Beckinson & Shionogi on asthma, COPD and chronic cough. **IMA**
485 received unrestricted research grants from the EU-IMI, GSK, Sanofi, MRC and EPSRC and lecture fees
486 from AZ, Eurodrug, Sanofi and Sunovion, and he also serves on advisory boards for Chiesi, GSK, Kineset

487 and Sanofi and fees for expert testimony from Chiesi. **RD** has no specific conflicts in respect of this study
488 but wishes, nevertheless, to disclose receiving payments for participation in advisory board meetings
489 with GSK, Regeneron and Kymab, and he is also co-founder and a shareholder of Synairgen, a University
490 of Southampton spin-out company where he has shares and is a consultant. **SSW**'s BioSci Consulting
491 company received consulting fees from Kings College Hospital NHS Foundation Trust, Academic Medical
492 Research, AMC Medical Research BV, Asthma UK, Athens Medical School, Boehringer Ingelheim
493 International GmbH, CHU de Toulouse, CIRO, DS Biologicals Ltd, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE
494 LAUSANNE, European Respiratory Society, FISEVI, Fluidic Analytics Ltd., Fraunhofer IGB, Fraunhofer
495 ITEM, GlaxoSmithKline Research & Dev Ltd, Holland & Knight, Karolinska Institutet Fakturor, KU Leuven,
496 Longfonds, National Heart & Lung Institute, Novartis Pharma AG, Owlstone Medical Limited, PExA AB,
497 UCB Biopharma S.P.R.L., Umeå University, Univ. Hosptial Southampton NHS Foundation Trust, Università
498 Campus Bio-Medico di Roma, Universita Cattolica Del Sacro Cuore, Universität Ulm, University of Bern,
499 University of Edinburgh, University of Hull, University of Leicester, University of Loughborough,
500 University of Luxembourg, University of Manchester, University of Notthingham, Vlaams Brabant, Dienst
501 Europa, Imperial College London, Boehringer Ingelheim, Breathomix, Gossamer Bio, Astrazeneca, CIBER,
502 OncoRadiomics, University of Leiden, University of Wurzburg, Chiesi Pharmaceutical, University of Liege,
503 Teva Pharmacauticals, Sanofi, Pulmonary Fibrosis Foundation, and Three Lakes Foundation. **KB** received
504 consulting fees from Sanofi and Astra Zeneca, lecture fees from Boehringer Ingelheim and fees for
505 participant in advisory board for ALK-Abelló Nordic. **SJS**'s institution received grants from Novo Nordic
506 Foundation (NNF) and Independent Research Fund Denmark's (DFF). **PJS** is scientific advisor and has an
507 officially non-substantial interest in the SME Breathomix BV that produces electronic noses and the
508 accompanying analysis algorithms. **GR**'s University received funding from EU IMI for this research. **HB**
509 reports receiving grants from The Lundbeck Foundation, The Ministry of Health, Danish Council for
510 Strategic Research, and The Capital Region Research Foundation. **AHM** has received grants from Health

511 Holland and she is the PI of a P4O2 (Precision Medicine for more Oxygen) public private partnership
512 sponsored by Health Holland involving many private partners that contribute in cash and/or in kind
513 (Boehringer Ingelheim, Breathomix, Fluidda, Ortec Logiqcare, Philips, Quantib-U, Smartfish, SODAQ,
514 Thirona, TopMD and Novartis), and received unrestricted research grants from GSK, Boehringer
515 Ingelheim and Vertex, and received consulting fees paid to her institution from Boehringer Ingelheim
516 and Astra Zeneca, and received honoraria for lectures paid to her institution from GlaxosmithKline;
517 outside the submitted work.

518 Besides the above mentioned, authors declare no potential, perceived, or real conflict of interest
519 regarding the content of this manuscript.

520

521

522 **References**

523 1. Bisgaard H, Nørgaard S, Sevelsted A, et al. Asthma-like symptoms in young children increase the
524 risk of COPD. J Allergy Clin Immunol. 2021;147(2):569-76.e9.
525 2. Pijnenburg MW, Frey U, De Jongste JC, et al. Childhood asthma: pathogenesis and phenotypes.
526 Eur Respir J. 2022;59(6).
527 3. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell.
528 2014;157(1):121-41.
529 4. Abdel-Aziz MI, Vijverberg SJH, Neerincx AH, et al. The crosstalk between microbiome and
530 asthma: Exploring associations and challenges. Clin Exp Allergy. 2019;49(8):1067-86.
531 5. Abdel-Aziz MI, Neerincx AH, Vijverberg SJ, et al. Omics for the future in asthma. Semin
532 Immunopathol. 2020;42(1):111-26.
533 6. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, et al. Sputum microbiome profiles identify severe
534 asthma phenotypes of relative stability at 12 to 18 months. J Allergy Clin Immunol. 2021;147(1):123-34.
535 7. Diver S, Richardson M, Haldar K, et al. Sputum microbiomic clustering in asthma and chronic
536 obstructive pulmonary disease reveals a Haemophilus-predominant subgroup. Allergy. 2020;75(4):808-
537 17.
538 8. Abdel-Aziz MI, Vijverberg SJH, Neerincx AH, et al. A multi-omics approach to delineate sputum
539 microbiome-associated asthma inflammatory phenotypes. Eur Respir J. 2022;59(1).
540 9. Charlson ES, Bittinger K, Haas AR, et al. Topographical continuity of bacterial populations in the
541 healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957-63.
542 10. Dickson RP, Erb-Downward JR, Freeman CM, et al. Spatial Variation in the Healthy Human Lung
543 Microbiome and the Adapted Island Model of Lung Biogeography. Ann Am Thorac Soc. 2015;12(6):821-
544 30.
545 11. Ahmed B, Cox MJ, Cuthbertson L, et al. Comparison of the upper and lower airway microbiota in
546 children with chronic lung diseases. PLoS One. 2018;13(8):e0201156.
547 12. Bisgaard H, Hermansen MN, Buchvald F, et al. Childhood asthma after bacterial colonization of
548 the airway in neonates. N Engl J Med. 2007;357(15):1487-95.
549 13. Thorsen J, Rasmussen MA, Waage J, et al. Infant airway microbiota and topical immune
550 perturbations in the origins of childhood asthma. Nat Commun. 2019;10(1):5001.
551 14. Raita Y, Camargo CA, Jr., Bochkov YA, et al. Integrated-omics endotyping of infants with
552 rhinovirus bronchiolitis and risk of childhood asthma. J Allergy Clin Immunol. 2021;147(6):2108-17.
553 15. Raita Y, Perez-Losada M, Freishtat RJ, et al. Nasopharyngeal metatranscriptome profiles of
554 infants with bronchiolitis and risk of childhood asthma: a multicentre prospective study. Eur Respir J.
555 2022;60(1).
556 16. Chiu CY, Chan YL, Tsai MH, et al. Cross-talk between airway and gut microbiome links to IgE
557 responses to house dust mites in childhood airway allergies. Sci Rep. 2020;10(1):13449.
558 17. Perez-Losada M, Authelet KJ, Hoptay CE, et al. Pediatric asthma comprises different phenotypic
559 clusters with unique nasal microbiotas. Microbiome. 2018;6(1):179.
560 18. Fleming L, Murray C, Bansal AT, et al. The burden of severe asthma in childhood and
561 adolescence: results from the paediatric U-BIOPRED cohorts. Eur Respir J. 2015;46(5):1322-33.
562 19. Thorsen J, Stokholm J, Rasmussen MA, et al. Asthma and Wheeze Severity and the
563 Oropharyngeal Microbiota in Children and Adolescents. Ann Am Thorac Soc. 2022.
564 20. Bel EH, Sousa A, Fleming L, et al. Diagnosis and definition of severe refractory asthma: an
565 international consensus statement from the Innovative Medicine Initiative (IMI). Thorax.
566 2011;66(10):910-7.

567 21. Bousquet J. Global initiative for asthma (GINA) and its objectives. Clin Exp Allergy. 2000;30 Suppl
568 1(s1):2-5.
569 22. de Onis M, Onyango AW, Borghi E, et al. Development of a WHO growth reference for school-
570 aged children and adolescents. Bull World Health Organ. 2007;85(9):660-7.
571 23. Siles RI, Hsieh FH. Allergy blood testing: A practical guide for clinicians. Cleve Clin J Med.
572 2011;78(9):585-92.
573 24. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, et al. eNose breath prints as a surrogate biomarker
574 for classifying patients with asthma by atopy. J Allergy Clin Immunol. 2020;146(5):1045-55.
575 25. Juniper EF, Guyatt GH, Feeny DH, et al. Measuring quality of life in children with asthma. Qual
576 Life Res. 1996;5(1):35-46.
577 26. Juniper EF, Guyatt GH, Feeny DH, et al. Measuring quality of life in the parents of children with
578 asthma. Qual Life Res. 1996;5(1):27-34.
579 27. Nathan RA, Sorkness CA, Kosinski M, et al. Development of the asthma control test: a survey for
580 assessing asthma control. J Allergy Clin Immunol. 2004;113(1):59-65.
581 28. Liu AH, Zeiger R, Sorkness C, et al. Development and cross-sectional validation of the Childhood
582 Asthma Control Test. J Allergy Clin Immunol. 2007;119(4):817-25.
583 29. Reddel HK, Taylor DR, Bateman ED, et al. An official American Thoracic Society/European
584 Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical
585 asthma trials and clinical practice. Am J Respir Crit Care Med. 2009;180(1):59-99.
586 30. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010 [Available
587 from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
588 31. Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools
589 and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
590 32. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
591 EMBnetjournal. 2011;17(1):3.
592 33. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from
593 Illumina amplicon data. Nat Methods. 2016;13(7):581-3.
594 34. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved
595 data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
596 35. Irizarry RA, Bolstad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data.
597 Nucleic Acids Res. 2003;31(4):e15.
598 36. Liberzon A, Birger C, Thorvaldsdottir H, et al. The Molecular Signatures Database (MSigDB)
599 hallmark gene set collection. Cell Syst. 2015;1(6):417-25.
600 37. Textor J, van der Zander B, Gilthorpe MS, et al. Robust causal inference using directed acyclic
601 graphs: the R package 'dagitty'. Int J Epidemiol. 2016;45(6):1887-94.
602 38. Mortensen MS, Brejnrod AD, Roggenbuck M, et al. The developing hypopharyngeal microbiota in
603 early life. Microbiome. 2016;4(1):70.
604 39. Stokholm J, Blaser MJ, Thorsen J, et al. Maturation of the gut microbiome and risk of asthma in
605 childhood. Nat Commun. 2018;9(1):141.
606 40. Bel EH. Clinical phenotypes of asthma. Curr Opin Pulm Med. 2004;10(1):44-50.
607 41. Howrylak JA, Fuhlbrigge AL, Strunk RC, et al. Classification of childhood asthma phenotypes and
608 long-term clinical responses to inhaled anti-inflammatory medications. J Allergy Clin Immunol.
609 2014;133(5):1289-300, 300 e1-12.
610 42. Lee JK, Han D. Atopic dermatitis is an important comorbidity in severe asthma. Ann Allergy
611 Asthma Immunol. 2018;120(6):661-2.
612 43. Jackson DJ, Sykes A, Mallia P, et al. Asthma exacerbations: origin, effect, and prevention. J Allergy
613 Clin Immunol. 2011;128(6):1165-74.

614 44. Murray CS, Poletti G, Kebadze T, et al. Study of modifiable risk factors for asthma exacerbations:
615 virus infection and allergen exposure increase the risk of asthma hospital admissions in children. Thorax.
616 2006;61(5):376-82.
617 45. Jackson DJ, Busse WW, Bacharier LB, et al. Association of respiratory allergy, asthma, and
618 expression of the SARS-CoV-2 receptor ACE2. J Allergy Clin Immunol. 2020;146(1):203-6 e3.
619 46. Thorsen J, Stokholm J, Rasmussen MA, et al. The Airway Microbiota Modulates Effect of
620 Azithromycin Treatment for Episodes of Recurrent Asthma-like Symptoms in Preschool Children: A
621 Randomized Clinical Trial. Am J Respir Crit Care Med. 2021;204(2):149-58.
622 47. Powell EA, Fontanella S, Boakes E, et al. Temporal association of the development of
623 oropharyngeal microbiota with early life wheeze in a population-based birth cohort. EBioMedicine.
624 2019;46:486-98.
625 48. Boutin S, Depner M, Stahl M, et al. Comparison of Oropharyngeal Microbiota from Children with
626 Asthma and Cystic Fibrosis. Mediators Inflamm. 2017;2017:5047403.
627 49. Rigauts C, Aizawa J, Taylor SL, et al. R othia mucilaginosa is an anti-inflammatory bacterium in
628 the respiratory tract of patients with chronic lung disease. Eur Respir J. 2022;59(5).
629 50. Torrego A, Hew M, Oates T, et al. Expression and activation of TGF-beta isoforms in acute
630 allergen-induced remodelling in asthma. Thorax. 2007;62(4):307-13.
631 51. Al-Alawi M, Hassan T, Chotirmall SH. Transforming growth factor beta and severe asthma: a
632 perfect storm. Respir Med. 2014;108(10):1409-23.
633 52. Sanjabi S, Oh SA, Li MO. Regulation of the Immune Response by TGF-beta: From Conception to
634 Autoimmunity and Infection. Cold Spring Harb Perspect Biol. 2017;9(6).
635 53. Kwak HJ, Park DW, Seo JY, et al. The Wnt/beta-catenin signaling pathway regulates the
636 development of airway remodeling in patients with asthma. Exp Mol Med. 2015;47(12):e198.
637 54. Jia XX, Zhu TT, Huang Y, et al. Wnt/beta-catenin signaling pathway regulates asthma airway
638 remodeling by influencing the expression of c-Myc and cyclin D1 via the p38 MAPK-dependent pathway.
639 Exp Ther Med. 2019;18(5):3431-8.
640 55. Sharma S, Tantisira K, Carey V, et al. A role for Wnt signaling genes in the pathogenesis of
641 impaired lung function in asthma. Am J Respir Crit Care Med. 2010;181(4):328-36.
642 56. Rogan MR, Patterson LL, Wang JY, et al. Bacterial Manipulation of Wnt Signaling: A Host-
643 Pathogen Tug-of-Wnt. Front Immunol. 2019;10:2390.
644 57. Hua K, Li Y, Zhou H, et al. Haemophilus parasuis Infection Disrupts Adherens Junctions and
645 Initializes EMT Dependent on Canonical Wnt/beta-Catenin Signaling Pathway. Front Cell Infect Microbiol.
646 2018;8:324.
647 58. Skronska-Wasek W, Durlanik S, Le HQ, et al. The antimicrobial peptide S100A8/A9 produced by
648 airway epithelium functions as a potent and direct regulator of macrophage phenotype and function. Eur
649 Respir J. 2022;59(4).
650 59. Chotirmall SH, Bogaert D, Chalmers JD, et al. Therapeutic Targeting of the Respiratory
651 Microbiome. Am J Respir Crit Care Med. 2022;206(5):535-44.
652 60. Leitao Filho FS, Takiguchi H, Akata K, et al. Effects of Inhaled Corticosteroid/Long-Acting β(2)-
653 Agonist Combination on the Airway Microbiome of Patients with Chronic Obstructive Pulmonary Disease:
654 A Randomized Controlled Clinical Trial (DISARM). Am J Respir Crit Care Med. 2021;204(10):1143-52.

655

656

657

658

659 **Tables:**

660 **Table 1:** Baseline characteristics for the recruited children with asthma or wheezing.

Characteristic	All children with microbiota profiles (n = 241)
Age in years, median (IQR)	6.0 (4.0, 13.0)
Female, n (%)	96/241 (40%)
Ethnicity, n (%)	
• Caucasian (white)	192/241 (80%)
• Black African	11/241 (5%)
• Central Asian	6/241 (2%)
• East Asian	2/241 (1%)
• South Asian	2/241 (1%)
• Arabic North Heritage	6/241 (2%)
• Multiple Races	18/241 (7%)
• Other	4/241 (2%)
Body mass index (BMI) z-score, median (IQR)	0.61 (0.01, 1.42) (n=240)
Asthma/wheezing cohort, n (%)	
• Cohort A (severe school-age asthma)	86/241 (36%)
• Cohort B (mild-to-moderate school-age asthma)	39/241 (16%)
• Cohort C (severe preschool wheezing)	65/241 (27%)
• Cohort D (mild-to-moderate preschool wheezing)	51/241 (21%)
Overall atopic sensitization[¶], n (%)	159/226 (70%)
Asthma control (well-controlled)[#], n (%)	72/181 (40%)
Quality of life (QOL) score-average, median (IQR)	5.30 (3.70, 6.47) (n=238)
Lung function test, median (IQR)	
• FEV_1 % predicted pre-salbutamol	92.9 (81.6, 106.7) (n=135)
• FEV_1 % predicted post-salbutamol	103.5 (92.6, 113.0) (n=143)
• FE_{NO} (in ppb)	30.0 (15.6, 57.0) (n=117)
Asthma medications, n (%)	
• ICS	212/241 (88%)
• SABA	227/241 (94%)

• LABA	137/241 (57%)
• OCS (maintenance)	40/241 (17%)
• SAMA	15/241 (6%)
• LAMA	3/241 (1%)
• LTRA	143/241 (59%)
• Nasal steroids	18/241 (7%)
Antibiotics intake, n (%)	
• Antibiotics (recent)	30/241 (12%)
• Antibiotics (recent & previous)	36/241 (15%)

Categorical variables are described as n (% of n), and continuous variables as median (interquartile range, (IQR)). BMI: body mass index, FEV_1: forced expiratory volume in 1 second, FE_{NO}: fractional exhaled nitric oxide, ICS: inhaled corticosteroids, LTRA: leukotriene antagonist, SABA: short-acting beta-agonist, SAMA: short-acting muscarinic antagonists, LABA: long-acting beta-agonist, LAMA: long-acting muscarinic antagonists, OCS: oral corticosteroids, ppb: part per billion, QOL: quality of life. For continuous measures (variables), the number of samples (n=) available for a specific measure is only provided when missing data were present. ¶ Overall atopic sensitization refers to atopic sensitization to aero- or food-allergens by either a positive skin prick test (wheal diameter ≥3mm) and/or a positive allergen-specific IgE (≥0.35kU/L) # A (childhood) asthma control test score >19 indicates well-controlled school-age asthma or preschool-age wheezing.

671 **Table 2:** Demographic and clinical characteristics of the oropharyngeal taxa-driven clusters.

Characteristic	Cluster 1 (n = 50)	Cluster 2 (n = 88)	Cluster 3 (n = 21)	Cluster 4 (n = 82)	*P*-value
Demographics					
Age in years, median (IQR)	5.0 (3.3-13.0)	9.5 (4.8-13.0)	5.0 (3.0-12.0)	5.0 (3.3-12.0)	0.144
Female, n (%)	21/50 (42%)	38/88 (43%)	5/21 (24%)	32/82 (39%)	0.430
Ethnicity, n (%)					0.427
• Caucasian (white)	44/50 (88%)	64/88 (73%)	19/21 (90%)	65/82 (79%)	
• Black African	0/50 (0%)	4/88 (5%)	1/21 (5%)	6/82 (7%)	
• Central Asian	2/50 (4%)	3/88 (3%)	0/21 (0%)	1/82 (1%)	
• East Asian	1/50 (2%)	0/88 (0%)	0/21 (0%)	1/82 (1%)	
• South Asian	0/50 (0%)	0/88 (0%)	0/21 (0%)	2/82 (2%)	
• Arabic North Heritage	1/50 (2%)	4/88 (5%)	0/21 (0%)	1/82 (1%)	
• Multiple Races	2/50 (4%)	11/88 (12%)	1/21 (5%)	4/82 (5%)	
• Other	0/50 (0%)	2/88 (2%)	0/21 (0%)	2/82 (2%)	
Body mass index (BMI) z-score, median (IQR)	0.59 (-0.12-1.01) (n=49)	0.77 (0.05-1.77) (n=88)	0.27 (0.10-1.08) (n=21)	0.61 (0.02-1.15) (n=82)	0.544
Delivery mode (C-section), n (%)	12/50 (24%)	25/88 (28%)	3/21 (14%)	18/82 (22%)	0.335
Smoking history, n (%)					
• Maternal smoking during pregnancy	5/50 (10%)	13/86 (15%)	2/20 (10%)	6/82 (7%)	0.453
• Second hand smoking	12/50 (24%)	18/85 (21%)	4/21 (19%)	11/79 (14%)	0.494
• Cotinine present (in urine)	4/43 (9%)	10/74 (14%)	2/20 (10%)	8/74 (11%)	0.911
Clinical characteristics					
Asthma/wheezing cohort, n (%)					0.057
• Cohort A (severe school-age asthma)	13/50 (26%)	39/88 (44%)	5/21 (24%)	29/82 (35%)	

• Cohort B (mild-to-moderate school-age asthma)	10/50 (20%)	16/88 (18%)	4/21 (19%)	9/82 (11%)	
• Cohort C (severe preschool wheezing)	15/50 (30%)	20/88 (23%)	10/21 (48%)	20/82 (24%)	
• Cohort D (mild-to-moderate preschool wheezing)	12/50 (24%)	13/88 (15%)	2/21 (10%)	24/82 (29%)	
History of diagnosed allergic disorders (ever), n (%)					
• **Atopic dermatitis (ever), n (%)**	**24/50 (48%)**	**68/87 (78%)**	**13/21 (62%)**	**55/81 (68%)**	**0.003**
• Allergic rhinitis (ever), n (%)	23/46 (50%)	42/82 (51%)	12/21 (57%)	34/78 (44%)	0.633
Asthma control (well-controlled)[#], n (%)	18/38 (47%)	24/72 (33%)	8/13 (62%)	22/58 (38%)	0.188
Quality of life (QOL) score-average, median (IQR)	5.4 (4.1-6.7) (n=50)	5.3 (3.3-6.4) (n=86)	5.3 (4.6-6.4) (n=21)	5.2 (3.7-6.5) (n=81)	0.853
Lung function test, median (IQR)					
• FEV_1 % predicted pre-salbutamol	97.5 (86.6-108.1) (n=28)	88.9 (74.3-99.31) (n=55)	96.3 (87.7-108.8) (n=11)	95.5 (87.0-106.9) (n=41)	0.108
• **FEV_1 % predicted post-salbutamol**	**108.7 (95.9-117.5) (n=29)**	**98.3 (88.8-107.7) (n=59)**	**107.2 (104.2-118.3) (n=10)**	**102.9 (94.1-109.9) (n=45)**	**0.047**
• FE_{NO} (in ppb)	43.0 (13.5-58.3) (n=20)	29.0 (16.0-50.0) (n=53)	35.0 (14.0-97.0) (n=9)	27.0 (14.5-58.0) (n=35)	0.954
Asthma/other mediations, n (%)					
• ICS	43/50 (86%)	80/88 (91%)	20/21 (95%)	69/82 (84%)	0.390
• SABA	48/50 (96%)	85/88 (97%)	19/21 (90%)	75/82 (91%)	0.422
• LABA	25/50 (50%)	56/88 (64%)	11/21 (52%)	45/82 (55%)	0.406
• OCS (maintenance)	6/50 (12%)	14/88 (16%)	5/21 (24%)	15/82 (18%)	0.623
• SAMA	4/50 (8%)	6/88 (7%)	0/21 (0%)	5/82 (6%)	0.670
• LAMA	1/50 (2%)	0/88 (0%)	0/21 (0%)	2/82 (2%)	0.503
• LTRA	24/50 (48%)	58/88 (66%)	13/21 (62%)	48/82 (59%)	0.235

• Nasal steroids	3/50 (6%)	9/88 (10%)	0/21 (0%)	6/82 (7%)	0.415
Antibiotics intake, n (%)					
• Antibiotics (recent)	7/50 (14%)	9/88 (10%)	2/21 (10%)	12/82 (15%)	0.803
• Antibiotics (recent & previous)	7/50 (14%)	10/88 (11%)	4/21 (19%)	15/82 (18%)	0.600

672 Categorical variables are described as n (% of n), and continuous variables as median (interquartile range, (IQR)). BMI: body mass index, FEV_1:
673 forced expiratory volume in 1 second, FE_{NO}: fractional exhaled nitric oxide, ICS: inhaled corticosteroids, LTRA: leukotriene antagonist, SABA:
674 short-acting beta-agonist, SAMA: short-acting muscarinic antagonists, LABA: long-acting beta-agonist, LAMA: long-acting muscarinic antagonists,
675 OCS: oral corticosteroids, ppb: part per billion, QOL: quality of life. For continuous measures (variables), the number of samples (n=) available for
676 a specific measure is only provided when missing data were present. P-values were calculated by a Pearson's Chi-square test with Monte Carlo
677 simulation (10000 permutations) or by a Kruskal-Wallis h test as appropriate. # A (childhood) asthma control test score >19 indicates well-
678 controlled school-age asthma or preschool-age wheezing.

679

680

681

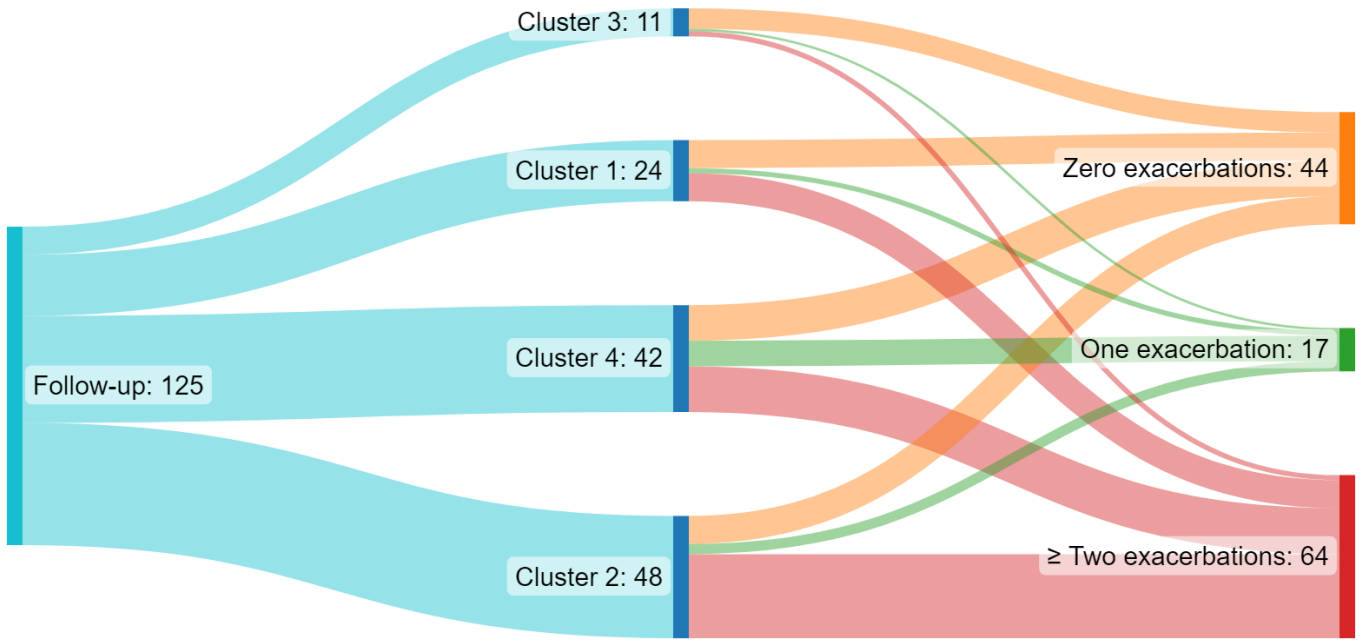
682 **Table 3:** Characteristics of atopic sensitization among the 4 taxa-driven clusters.

Characteristic	Cluster 1 (n = 50)	Cluster 2 (n = 88)	Cluster 3 (n = 21)	Cluster 4 (n = 82)	*P*-value
Overall atopic sensitization¶, n (%)	**25/47 (53%)**	**65/83 (78%)**	**13/18 (72%)**	**56/78 (72%)**	**0.022**
Common aeroallergens sensitization, n (%)	**25/47 (53%)**	**62/83 (75%)**	**13/18 (72%)**	**46/78 (59%)**	**0.045**
Common aeroallergens, n (%)					
• House dust mites	19/46 (41%)	46/81 (57%)	11/17 (65%)	31/77 (40%)	0.068
• Tree	17/46 (37%)	33/79 (42%)	6/16 (38%)	21/70 (30%)	0.522
• Mold	6/43 (14%)	27/78 (35%)	4/16 (25%)	19/71 (27%)	0.111
• **Grass**	**13/46 (28%)**	**45/81 (56%)**	**8/18 (44%)**	**29/74 (39%)**	**0.021**
• Pet animal (cat/dog)	22/46 (48%)	52/81 (64%)	7/18 (39%)	36/75 (48%)	0.079
Number of aeroallergen classes, median (IQR)	**0.0 (0.0-3.0) (n=47)**	**3.0 (1.0-4.00) (n=83)**	**1.5 (0.0-3.8) (n=18)**	**1.0 (0.0-3.0) (n=78)**	**0.021**
Total IgE (in kU/L)., median (IQR)	292.6 (54.7-924.7) (n=38)	245.1 (95.8-1,002.5) (n=76)	105.0 (52.8-310.0) (n=17)	150.0 (38.9-682.3) (n=68)	0.240

683 Categorical variables are described as n (% of n), and continuous variables as median (interquartile range, (IQR)). ¶ Overall atopic sensitization
684 refers to sensitization to either aero- and/or food-allergens by either a positive skin prick test (wheal diameter ≥3mm) and/or a positive allergen-
685 specific IgE (≥0.35kU/L). P-values were calculated by a Pearson's Chi-square test with Monte Carlo simulation (10000 permutations) or by a
686 Kruskal-Wallis H test as appropriate.

687 **Table 4:** Baseline microbiota cluster assignments and the frequency of exacerbations per year during follow-up (12-18 months after baseline
688 visit).

Characteristic	Cluster 1 (n = 24)	Cluster 2 (n = 48)	Cluster 3 (n = 11)	Cluster 4 (n = 42)	*P*-value
Frequency of exacerbations/year during follow-up, n (%)					**0.005**
• **Zero**	**11/24 (46%)**	**11/48 (23%)**	**8/11 (73%)**	**14/42 (33%)**	
• **One**	**2/24 (8%)**	**4/48 (8%)**	**1/11 (9%)**	**10/42 (24%)**	
• **≥ Two**	**11/24 (46%)**	**33/48 (69%)**	**2/11 (18%)**	**18/42 (43%)**	



689 Lower panel shows Sankey diagram for Baseline microbiota cluster assignments and frequency of exacerbations/year during follow-up.

Box 1: Summary overview of the clinical characteristics of the taxa-driven clusters.

Cluster 1

Cluster 1 included 50 (20.7%) patients. Cluster 1 had the least children with atopic dermatitis diagnosis (n=24/50, 48%, post-hoc $p = 0.001$), the least children with atopic sensitization (n=25/47, 53%, post-hoc $p = 0.003$) particularly to grass pollens (n=13/46, 28%, post-hoc $p = 0.020$), and had the lowest median number of the aeroallergens sensitizations (median = 0, IQR (0-3)) compared with the other clusters (all *p*s < 0.05). Generally, patients within Cluster 1 showed the lowest median ESs of blood transcriptomic signatures related to the Wnt/β-Catenin (p <0.001, q <0.01), the TGF-β signaling (p <0.01, q <0.05) and the myogenesis (p <0.05, q >0.05) MSigDB hallmark pathways.

Cluster 2

Cluster 2 was the largest cluster and included 88 patients (36.5%). Cluster 2 included older school-aged asthmatics (median age=9.5, IQR (4.8-13.0)), and exhibited the highest percentage of school-age severe asthmatics (n=39/88, 44%, post-hoc $p = 0.034$) compared with the other clusters (median age = 5), however; neither differences were statistically significant at overall clusters' level ($p = 0.144$ and 0.057, respectively). Cluster 2 showed the highest percentage of atopic dermatitis diagnosis (n=68/87, 78%, post-hoc $p = 0.005$), the highest percentage of atopic sensitization to aeroallergens particularly grass pollens (n=45/81, 56%, post-hoc $p = 0.005$), and had highest median number of the aeroallergens sensitizations (median = 3, IQR (1-4)) compared with the other clusters. In addition, Cluster 2 had the lowest median values of FEV_1 % predicted values post-bronchodilator (median=98.3, IQR (88.8-107.7)) and had the highest percentage of children with severe symptoms who experienced future exacerbations (n=37/48, 77%), particularly ≥2 future exacerbations/year (n=33/48, 69%, post-hoc $p = 0.002$) during the follow-up period (all *p*s < 0.05). This Cluster showed the relatively highest median ESs of TGF-β signaling MSigDB hallmark pathway (p <0.01, q <0.05).

Cluster 3

Cluster 3 was the smallest cluster and included 21 patients (8.7%). This cluster had the highest percentage of preschoolers with severe wheezing (n=10/21, 48%, post-hoc $p = 0.026$) compared with the other clusters, however; this difference was not statically significant at the overall clusters' level ($p = 0.057$). This cluster had a significantly higher median values of FEV_1 % predicted post-bronchodilator compared with Cluster 2 (median 107.2, IQR (104.2-118.3), and 98.3 IQR (88.8-107.7), respectively, p <0.05), and had the highest percentage of children with severe symptoms (n=8/11, 73%, post-hoc $p = 0.006$) who did not experience any future exacerbations during the follow-up period.

Cluster 4

Cluster 4 was the second-largest cluster and included 82 patients (34%). This cluster included the second-highest percentage of children with severe symptoms (n=28/42, 67%) who experienced future exacerbations and particularly included the highest percentage of children with severe symptoms who experienced one exacerbation/year (n=18/42, 24%, post-hoc $p = 0.018$) during the follow-up period compared with the other clusters ($p < 0.05$). This cluster showed the relatively highest median ESs of the Wnt/β-Catenin signaling and second-highest median ESs of the TGF-β signaling MSigDB hallmark pathways (all *ps* <0.01, q <0.05).

690

691

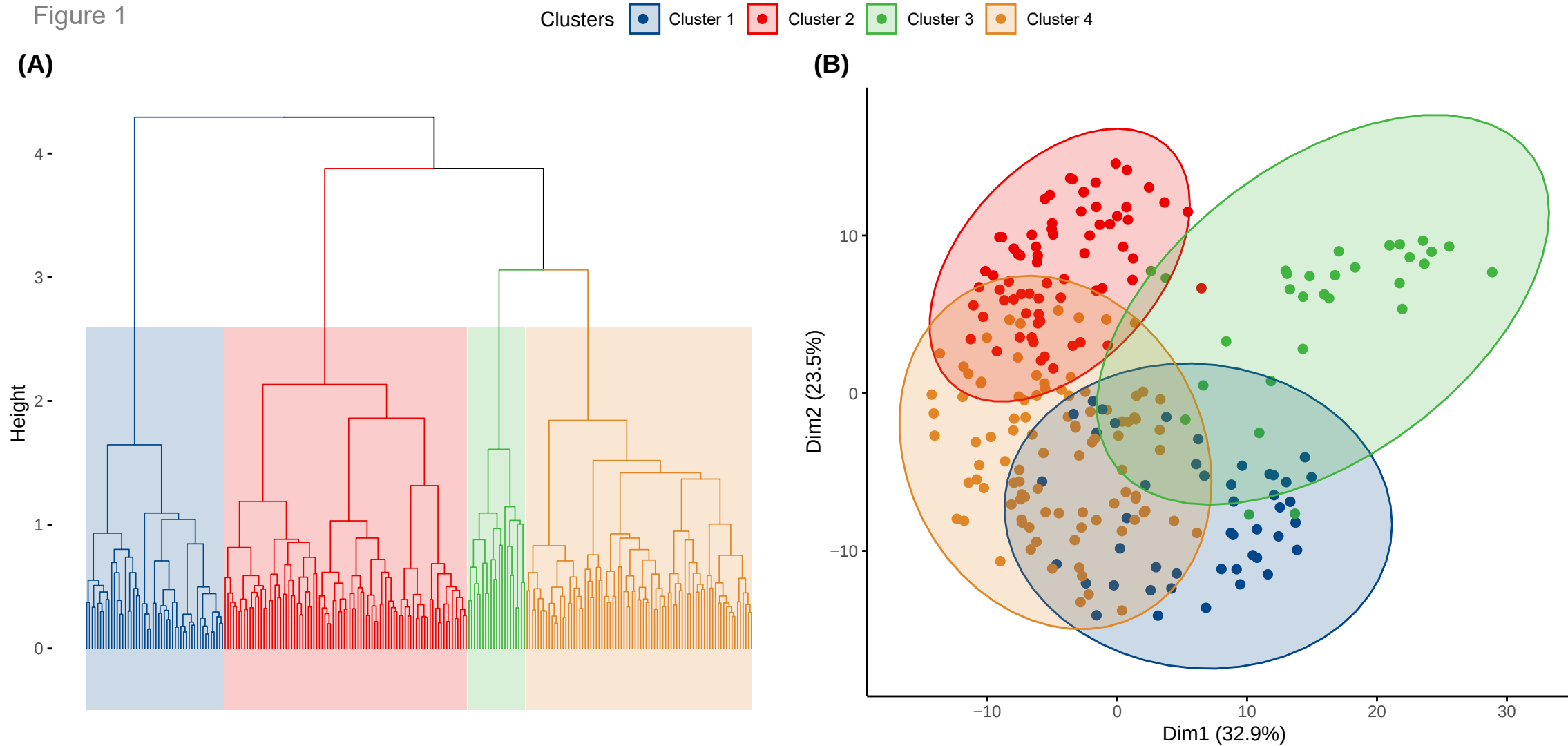
692 **Figure Legends:**

693 **Figure 1: A;** Hierarchical agglomerative cluster dendrogram of the Bray-Curtis dissimilarities showing the
694 4 main children groups. **B;** Principal component plot of the clustering by partition around the medoids
695 (PAM) shows 4 relatively detached 95% reference intervals ellipses of 4 children groups. The similarity in
696 participants' assignment between the two clustering algorithms was assessed by Rand index (RI= 0.85)
697 and Pearson Chi-Square test (χ^2= 498.1, $p < 1 \times 10^{-4}$) showing high similarity in the clustering
698 assignments. For assessing clusters stability, replacement of points by noise schemes (1000 iterations)
699 resulted in mean Jaccard similarity indices ranging from 0.74-1 for the 4 clusters by either hierarchical
700 clustering or PAM suggesting relatively stable clusters.

701 **Figure 2:** Differentially abundant bacterial taxa between the taxa-driven clusters using the Linear
702 discriminant analysis Effect Size (LEfSe) statistical method. The highest enriched bacterial taxa related to
703 each cluster are shown at phylum (**A**) and species (**B**) levels. Only differentially abundant bacterial
704 species with an FDR *q*-value <0.01 and an absolute LDA score >2 are depicted.

705 **Figure 3:** Box-and-whisker plots showing the enrichment scores of the Molecular Signatures Database
706 (MSigDB) gene sets between the taxa-driven clusters. Only the statistically significant gene sets are
707 shown, and those passing the multiple testing threshold (*q*-value <0.05) are shown in bold-face. Overall
708 *P*-values were generated by the Kruskal-Wallis H test and pairwise *P*-values were generated by Dunn's
709 post-hoc test.

Figure 1


Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
(A)
Height
4
3
2
1
0
(B)
Dim2 (23.5%)
10
0
−10
Dim1 (32.9%)
−10
0
10
20
30

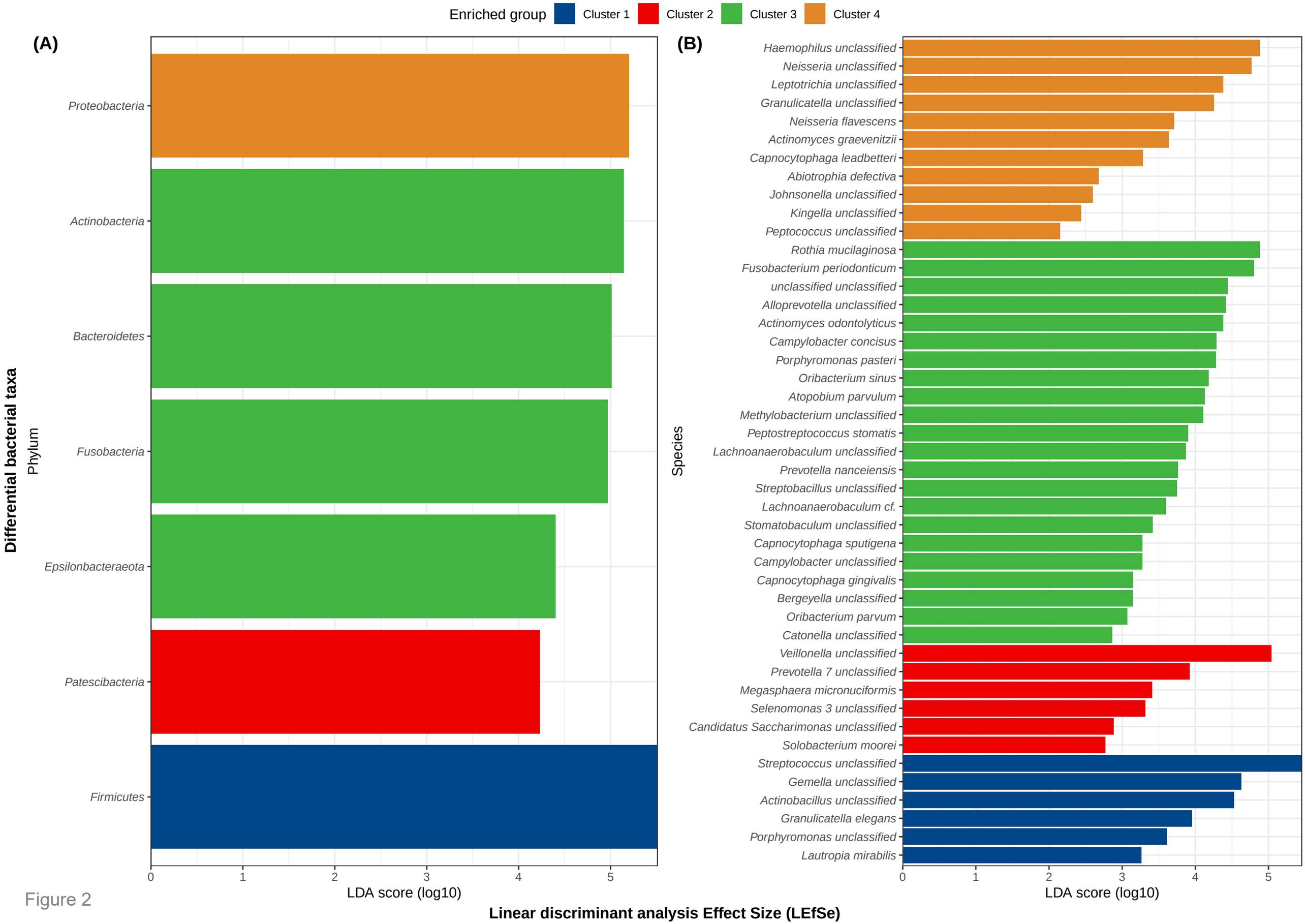
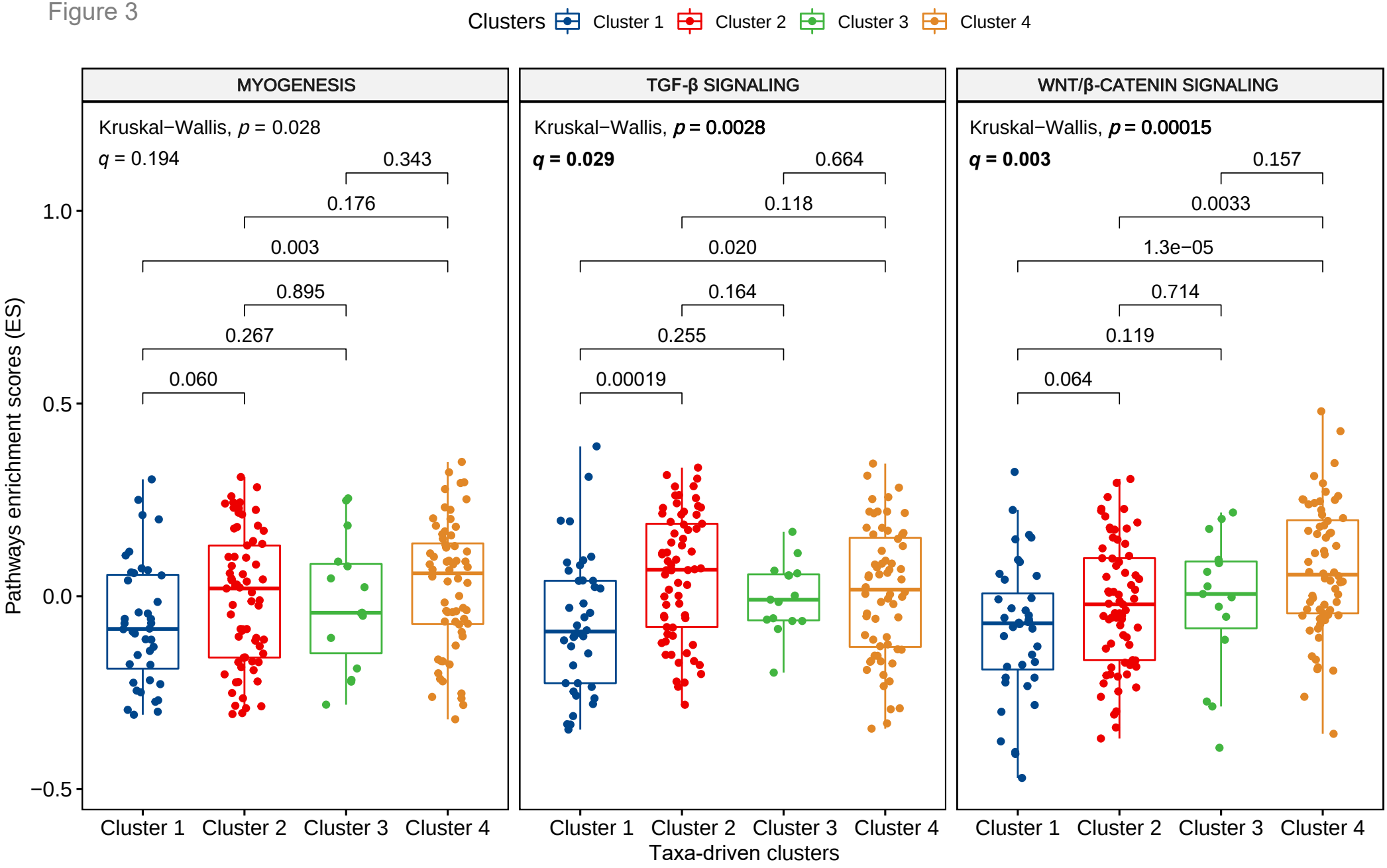

Enriched group
Cluster 1
Cluster 2
Cluster 3
Cluster 4
(A)
Differential bacterial taxa
Phylum
Proteobacteria
Actinobacteria
Bacteroidetes
Fusobacteria
Epsilonbacteraeota
Patescibacteria
Firmicutes
0
1
2
3
4
5
LDA score (log10)
(B)
Species
Haemophilus unclassified
Neisseria unclassified
Leptotrichia unclassified
Granulicatella unclassified
Neisseria flavescens
Actinomyces graevenitzii
Capnocytophaga leadbetteri
Abiotrophia defectiva
Johnsonella unclassified
Kingella unclassified
Peptococcus unclassified
Rothia mucilaginosa
Fusobacterium periodonticum
unclassified unclassified
Alloprevotella unclassified
Actinomyces odontolyticus
Campylobacter concisus
Porphyromonas pasteri
Oribacterium sinus
Atopobium parvulum
Methylobacterium unclassified
Peptostreptococcus stomatis
Lachnoanaerobaculum unclassified
Prevotella nanceiensis
Streptobacillus unclassified
Lachnoanaerobaculum cf.
Stomatobaculum unclassified
Capnocytophaga sputigena
Campylobacter unclassified
Capnocytophaga gingivalis
Bergeyella unclassified
Oribacterium parvum
Catonella unclassified
Veillonella unclassified
Prevotella 7 unclassified
Megasphaera micronuciformis
Selenomonas 3 unclassified
Candidatus Saccharimonas unclassified
Solobacterium moorei
Streptococcus unclassified
Gemella unclassified
Actinobacillus unclassified
Granulicatella elegans
Porphyromonas unclassified
Lautropia mirabilis
0
1
2
3
4
5
LDA score (log10)
Linear discriminant analysis Effect Size (LEfSe)

Figure 2

Figure 3


Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
MYOGENESIS
Kruskal−Wallis, p = 0.028
q = 0.194
0.343
0.176
0.003
0.895
0.267
0.060
TGF-β SIGNALING
Kruskal−Wallis, p = 0.0028
q = 0.029
0.664
0.118
0.020
0.164
0.255
0.00019
WNT/β-CATENIN SIGNALING
Kruskal−Wallis, p = 0.00015
q = 0.003
0.157
0.0033
1.3e−05
0.714
0.119
0.064
Pathways enrichment scores (ES)
1.0
0.5
0.0
−0.5
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Taxa-driven clusters

Online data supplement

Oropharyngeal microbiota clusters in children with asthma/wheeze associate with allergy, blood transcriptomic immune pathways and exacerbations risk

Mahmoud I. Abdel-Aziz, PhD[1,2,3,4]; Jonathan Thorsen, MD, PhD[5,6]; Simone Hashimoto, MD, PhD[1,2,3,7]; Susanne J. H. Vijverberg, PhD[1,2,3]; Anne H. Neerincx, PhD[1,2,3]; Paul Brinkman, PhD[1,2,3]; Wim van Aalderen, MD, PhD[7]; Jakob Stokholm, MD, PhD[5,8]; Morten Arendt Rasmussen, MD, PhD[5,8]; Michael Roggenbuck-Wedemeyer, MD, PhD[9,10]; Nadja H. Vissing, MD, PhD[5]; Martin Steen Mortensen, PhD[9]; Asker Daniel Brejnrod, PhD[11]; Louise J. Fleming, MD[12]; Clare S. Murray, MD[13]; Stephen J. Fowler, MD, PhD[13]; Urs Frey, MD, PhD[14]; Andrew Bush, MD, FRCP, FRCPCH[12]; Florian Singer, MD, PhD[15]; Gunilla Hedlin, MD, PhD[16]; Björn Nordlund, RN, PhD[16]; Dominick E. Shaw, PhD[17]; Kian Fan Chung, MD, DSc[12]; Ian M. Adcock, PhD[12]; Ratko Djukanovic, MD, PhD[18]; Charles Auffray, PhD[19]; Aruna T. Bansal, PhD[20]; Ana R. Sousa, PhD[21]; Scott S. Wagers, MD, PhD[22]; Bo Lund Chawes, MD, PhD[6]; Klaus Bønnelykke, MD, PhD[6]; Søren Johannes Sørensen, PhD[9]; Aletta D. Kraneveld PhD[23]; Peter J. Sterk, MD, PhD[1,2]; Graham Roberts, DM[18]; Hans Bisgaard, MD, PhD[5]; Anke H. Maitland-van der Zee, PharmD, PhD[1,,2,3,7]; on behalf of the U-BIOPRED Study Group.

1 Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands.

2 Amsterdam Institute for Infection and Immunity, Amsterdam, the Netherlands.

3 Amsterdam Public Health, Amsterdam, the Netherlands.

4 Department of Clinical Pharmacy, Faculty of Pharmacy, Assiut University, 71526 Assiut, Egypt.

5 COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

6 Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

24 7 Department of Paediatric Pulmonary Medicine, Emma Children's Hospital, Amsterdam UMC, University of
25 Amsterdam, 1105 AZ Amsterdam, the Netherlands.
26 8 Department of Food Science, University of Copenhagen, Frederiksberg, Denmark.
27 9 Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
28 10 Novozymes, Bagsvaerd, Denmark.
29 11 Section of Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby,
30 Denmark.
31 12 National Heart and Lung Institute, Imperial College London, and Royal Brompton and Harefield NHS Trust,
32 London, United Kingdom.
33 13 Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology,
34 Medicine and Health, University of Manchester, and Manchester Academic Health Science Centre and NIHR
35 Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom.
36 14 University Children's Hospital Basel, University of Basel, Basel, Switzerland.
37 15 Division of Paediatric Pulmonology and Allergology, Department of Paediatrics and Adolescent Medicine,
38 Medical University of Graz, Austria & Division of Paediatric Respiratory Medicine and Allergology, Department of
39 Paediatrics, Inselspital, Bern University Hospital, University of Bern, Switzerland.
40 16 Astrid Lindgren Children's Hospital, Karolinska University Hospital, and Department of Women's and Children's
41 Health, Karolinska Institutet, Stockholm, Sweden.
42 17 NIHR Respiratory Biomedical Research Unit, University of Nottingham, Nottingham, United Kingdom.18 NIHR
43 Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust and Clinical and
44 Experimental Sciences and Human Development and Health, University of Southampton, Southampton, United
45 Kingdom.
46 19 European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon, France.
47 20 Acclarogen Ltd, St John’s Innovation Centre, Cambridge, United Kingdom.
48 21 Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park, United Kingdom.
49 22 BioSci Consulting, Maasmechelen, Belgium.

50 23 Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University,

51 Utrecht, the Netherlands.

52 **Atopy**

53 Atopy was defined as a positive skin prick test (SPT) defined by a wheal diameter ≥3mm and/or a positive
54 allergen-specific IgE ≥0.35kU/L to a pre- specified allergen list. Skin prick testing was done using single
55 headed lancet and positive (histamine 10mg/ml) and negative (saline) controls (ALK-Abello, Horshølm,
56 Denmark) as previously described [1]. All sites tested to the most common aeroallergens, namely house
57 dust mite (mixture of *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*), animals (cat and
58 dog), grass pollen mixture, tree pollen mixture and mold (aspergillus). Up to three additional allergens,
59 such as cockroach and *Olea europea* and *Parietaria* were performed where relevant to a specific centre.
60 Allergen-specific IgE testing was measured for the above-listed most common aeroallergens (Thermo
61 Fisher, Uppsala, Sweden).

62 **Samples collection and omics analysis**

63 **Oropharyngeal swabs and 16S gene V4 microbiome sequencing**

64 A single oropharyngeal (throat) swab was collected from the posterior pharyngeal wall, behind the uvula,
65 of each subject by research assistants using dry flocked swabs (ESwab, BD, NJ,USA). The swabs were
66 placed in a Liquid Amies transport medium and stored at -20°C for further processing. Genomic
67 FastDNA™ SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA) was used for DNA extraction and DNA
68 extracts were checked by gel electrophoresis for consistency and quality. A two-step PCR process was
69 implemented to amplify the 16S ribosomal RNA (rRNA) by targeting the hypervariable V4 region by
70 primers (515F: 5'-GTGCCAGCMGCCGCGGTAA and 806R: 5'-GGACTACHVGGGTWTCTAAT) and
71 normalization for amplicon contents was performed before undergoing sequencing. Sequencing was
72 done using the Illumina MiSeq Desktop Sequencer (Illumina Inc., CA, USA) as described previously [2].
73 Paired-end reads (250 bp x 2) were pooled for subsequent quality control and bioinformatics processing.
74 The quality of the sequence reads were checked using FastQC [3] and MultiQC [4]. Primers were

removed using Cutadapt [5]. The dada2 pipeline was used to quantify the Amplicon Sequence Variants (ASVs) as described in detail previously [6]. Briefly, the pipeline works by performing quality filtering and trimming, dereplicating sequences, learning dataset-specific error rates, denoising by removing potentially containing errors sequences, merging paired-end reads while removing mismatches to reduce errors, constructing amplicon sequence variants (ASVs), removing chimera, and running taxonomic classification of ASVs using different publicly available databases. During the trimming and filtering step of the forward and reverse sequencing reads, quartiles of the quality score distribution of nucleotide positions were inspected. Parameters for the reads filtering were used as follows; truncQ=2 (truncate reads at the first instance of a quality score less than or equal to 2), maxEE=5 (reads with higher than 5 expected errors will be discarded), and rm.phix=TRUE (discard reads that match against the phiX genom). Then, learning of the error rates was performed. The estimated error rates were in a good fit to the observed rates, and the error rates dropped with increased quality as expected [6]. Afterwards, the, forward and reverse reads were merged using the default parameters in dada2 (minimum length of overlap=20 nucleotides, and maximum nucleotides mismatches=0). Subsequently, chimeric sequences were removed by implementing the Bimera method (removeBimeraDenovo function in dada2). Fourteen % of the total sequences were identified as chimera and subsequently removed from the final identified ASVs table. Finally, taxonomy of the identified ASVs was annotated using the Ribosomal Database Project (RDP) Naive Bayesian Classifier algorithm [7] implemented in dada2 using the default parameters (minimum bootstrap confidence for assigning a taxonomic level=50) against the Silva database version 132 [8]. Negative extraction (sterile water) and positive (mock community, BEI Resources, NIAID, NIH) controls were added to each plate (sequencing run) as previously described [2], and checked against real samples by comparing Shannon α-diversity (using Mann-Whitney U test) and by relative abundance plots. As a sanity check, a statistical method to remove potentially contaminant sequences (decontam R

98 package) [9] was used to filter the ASVs table (using the prevalence method) before downstream
99 statistical analysis.

100 **Blood collection and transcriptomics analysis**

101 Peripheral blood samples were collected and RNA was isolated using the PAXgene Blood RNA kit
102 (PreAnalytiX, Hombrechtikon, Switzerland) including on-column DNase treatment (QIAGEN, Valencia, CA)
103 and stored for further processing. A 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) was used to
104 assess RNA integrity, by which samples with an RNA integrity number ≥6 were further processed for
105 gene expression analysis. Microarray gene expression was done by hybridization onto Affymetrix HT HG-
106 U133+ PM arrays according to Affymetrix technical protocols using a GeneTitan Instrument (Affymetrix,
107 Santa Clara, CA). The microarray data are deposited in the Gene Expression Omnibus database (accession
108 number: GSE123750).

109 Raw microarray data (as .CEL files) were quality assessed and pre-processed by robust multi-array
110 average normalization (RMA) method (10) as implemented in the affy R package. Probes of low
111 expression were filtered by robust multi-array signal analysis for values<5 and also for batch/technical
112 effects. The intensity of the raw probe sets were background corrected and quantile-normalized by the
113 robust RMA (rma() function in the affy R package, using the following parameters normalize = TRUE and
114 background = TRUE), and log base 2 transformed, followed by z-scoring the values.

115 **Data analysis**

116 The general data analysis workflow is shown in Figure E2.

117 ***Clustering protocol***

118 A benchmarking clustering strategy was performed based on the approach utilized previously [10]. The
119 Bray-Curtis β-diversity dissimilarity measure was computed on the relative abundances of the amplicon

sequence variants (ASVs) to assess the within-participants overall variability in the microbiota profiles. An agglomerative hierarchical Ward2 clustering algorithm was applied [11] and the optimum number of clusters was determined based on the majority vote of 4 statistical measures, namely; average silhouettes width [12], the total within-cluster sum of square (WCSS) [13], Calinski-Harabasz [14], and Hubert and Levin C [15] indices. Cluster assignment of the participants was internally validated by using partition around the medoids (PAM) [16]. Agreement in the clustering of participants' assignments between hierarchical Ward, and PAM clustering was quantified by means of Pearson Chi-Square test with Monte Carlo simulation (10000 permutations) and Rand index (ranges from 0 to 1, indicating no agreement to high agreement, respectively) [17]. The clusters were visualized in two-dimensional space by the non-metric multidimensional scaling (MDS) and principal co-ordinate analysis (PCoA) on the Bray-Curtis dissimilarity measure, in which the 95% confidence ellipses were depicted. Clustering was further validated using topological data analysis (TDA) mapper tmap [18]. In tmap-TDA analysis, the density-based spatial clustering algorithm (DBSCAN) was utilized (optimum epsilon was calculated using the optimize_dbscan_eps function in tmap, resolution was selected=20 and % overlap=65, and lenses were computed using MDS and PCoA analysis on Bray-Curtis dissimilarity measure). The spatial analysis of functional enrichment (SAFE) scores were calculated to quantify the enrichment level of the clusters' assignments of samples belonging to the subnetwork centered around the node (SAFE scores were calculated by network permutation of 5000 times) as described in detail previously [18]. The nodes in the generated TDA network were then colored according to the SAFE scores of the clusters' assignments. Cluster-wise stability of the hierarchical and PAM clusters was evaluated by replacement of points by noise (1000 iterations) with subsequent calculation of the Jaccard similarity index [19].

142 ***Statistical analysis***

143 The differential microbiota profiles shaping the revealed clusters were evaluated by looking at relative
144 abundances plots of the microbiota. The Linear discriminant analysis Effect Size (LEfSe) was used to
145 reveal the main differentially abundant taxa between the clusters at phylum and species levels [20] after
146 adjusting by multiple testing by Benjamini Hochberg false discovery rate (FDR) [21], where an absolute
147 linear discriminant analysis (LDA) score >2 and FDR *q*-value <0.01 were used as cut-offs for significance.
148 Similarly, the top enriched taxa by the SAFE algorithm in TDA analysis were depicted in the TDA map to
149 check concordance with the LEfSe method revealed top differential bacterial taxa. Finally, to reveal
150 whether a minimal set of the differential bacterial taxa could be used to guide the participants'
151 stratification, an inference decision tree was built on the bacterial relative abundances to predict the
152 clusters assignment. The minimum set of taxa was determined based on the LEfSe method, by selecting
153 the top enriched taxa in each cluster (based on highest LDA score), resulting in 4 main bacterial genera
154 that were used as input for the inference decision tree. The data were randomly divided into 75:25%,
155 training and validation sets, respectively. A five-fold cross-validated model (using ctree2 method) was
156 built on the training dataset (hyper-parameters tuning was based on a grid; tuneLength=100,
157 maxdepth=1:50, and mincriterion=0.95 and 0.99). The performance of the model was tested on the left-
158 out validation set by calculating accuracy (95% CI).

159 Patient cluster distribution according to the inclusion center and season of sample collection
160 (metrological system) and differences in participants' demographic and clinical characteristics between
161 the revealed clusters were compared using Pearson Chi-Square test with Monte Carlo simulation (10000
162 permutations) or Kruskal Wallis H test tests as appropriate. Results are considered significant at alpha
163 level < 0.05. If the overall p-value of a Kruskal-Wallis H test was significant (<0.05), a post-hoc Dunn's test
164 was performed. If the overall p-value of a Pearson Chi-Square test was significant (<0.05), a post-hoc

165 analysis based on the residuals was done as previously described [22]. In addition, logistic and linear
166 regression models were used to investigate whether the clinical characteristics of interest are dependent
167 on the participants' clusters assignments after adjusting for multiple covariates, including age, sex,
168 ethnicity, BMI z-scored, center and season of inclusion, inhaled, oral and nasal corticosteroids intake,
169 reported recent antibiotic intake and potential technical covariates (library size, extraction plate ID; 2
170 plates, and sequencing run; 2 runs). As a first step, all covariates were forced into the regression models
171 and the microbiota-clusters were checked whether they are significantly associated with the clinical
172 characteristics of interest after adjusting for all other covariates. If the microbiota clusters were found to
173 be significantly associated with clinical characteristics of interest after adjusting for all covariates, a
174 stepwise (backward and forward selection) variable selection process [23] was then implemented in the
175 regression models to reduce the overall model error rate. All covariates (including microbiota-clusters)
176 had similar chance to be selected in the stepwise process. Multiple testing correction for the covariates
177 *P*-values was performed by FDR. Similarly, the cluster assignments of severe children with asthma or
178 wheezing were checked against the frequency of exacerbations during follow-up (using Chis-Square test
179 with Monte Carlo simulation, and with logistic regression after adjusting for covariates). The latter was
180 also depicted visually using a Sankey diagram. A simple mediation analysis was conducted to investigate
181 whether microbiota Cluster 2 is an independent predictor of future asthma exacerbations (≥2) or if this is
182 mediated by the atopic dermatitis status. For the mediation analysis, the outcome variable was >=2
183 exacerbations per year during follow-up, the predictor variable was the microbiota Cluster 2 (reference
184 level was the other clusters), and the mediator variable was atopic dermatitis status. A logistic regression
185 model was conducted to estimate association of the predictor variable into the mediator. A second
186 logistic regression model was conducted to estimate the association of the mediator with the outcome
187 variable in the presence of the predictor variable. Finally, the causal mediation was performed between
188 the two models using the mediate() function from the mediation R package (using default parameters),

where the 95% CIs for the estimates were computed by Quasi-Bayesian approximation (1000 simulations). Results of the mediation analysis were reported before and after adjusting for covariates.

Transcriptomic pathway enrichment analysis of the clusters was performed by an unsupervised approach using the gene set variation analysis (GSVA) based on the genes z-scores of the participants against a list of 21 Hallmark gene sets from the Molecular Signatures Database (MSigDB) release version 7.4 [24] that were previously reported to be associated with asthma disease (Table E1). The GSVA R package was used for this analysis (gsva function was utilized using the following parameters, method="gsva", abs.ranking=FALSE, min.sz=10, max.sz=500, rnaseq=FALSE and other parameters by default). The GSVA enrichment scores (ES) of the tested gene sets were compared between the clusters by Kruskal-Wallis H-test and multiple testing was corrected by FDR. Permutation analysis of the significant gene sets (pathways) was performed, in which the gene sets were resampled (10000 times) with random genes of similar size as the initially investigated Hallmark gene sets (i.e. randomly assigning genes to gene sets of similar size n as the hallmark gene sets). Enrichment scores from the randomly generated gene sets were checked for differences among the clusters using the Kruskal Wallis H test. The permutation analysis p-values were calculated using the following equation

$$\frac{sum\ (Kruskal{-}Wallis\ test\ statistic\ of\ the\ randomly\ generated\ gene\ sets \geq the\ test\ statistic\ of\ the\ original\ hallmark\ gene\ set)+1}{number\ of\ permutations+1}$$

. This permutation analysis was performed to check the null hypothesis that the significance is due to the presence of non-specific/non-relevant genes randomly present in a gene set. Linear regression models were used to investigate whether the enrichment scores of the significant gene sets are dependent on the participants' clusters assignments after adjusting for multiple covariates, including age, sex, ethnicity, BMI z-scored, site and season of inclusion, atopic sensitization, inhaled, oral and nasal corticosteroids intake, reported recent antibiotic intake and potential technical covariates (library size, extraction plate ID; 2 plates, and sequencing run; 2 runs). As a first step, all covariates were forced into the regression models and the microbiota-clusters were checked whether they are significantly associated with the

transcriptomic pathways of interest after adjusting for all other covariates. If the microbiota clusters were found to be significantly associated with the transcriptomic pathways of interest after adjusting for all covariates, a stepwise (backward and forward selection) variable selection process [23] was them implemented in the regression models to reduce the overall model error rate. All covariates (including microbiota clusters) had similar chance to be selected in the stepwise process. Multiple testing correction for the covariates *P*-values was performed by FDR. Assumptions of the linear regression models were checked using the Global Validation of Linear Model Assumptions as previously described [25].

All analyses were performed using R studio (version 1.1.453) with R software (version 3.6.1) supported with the following packages: phyloseq (v1.34.0), vegan (v2.6.2), stats (v4.0.3), MASS (v7.3.53), cluster (v2.1.0), factoextra (v1.0.7), fpc (v2.2.9), fossil (v0.4.0), fmsb (v0.7.3), performance (v0.9.1), ResourceSelection (v0.3.5), party (v1.3.9), gvlma (v1.0.0.3), GSVA (v1.40.1), microbiomeMarker (v0.0.1.9000), gtsummary (v1.5.0.9004), qwraps2 (v0.5.2), chisq.posthoc.test (v0.1.2), affy (v1.70.0), mediation (v4.5.0), ggplot2 (v3.3.6), and ggpubr (v0.4.0). TDA was performed in Python (version 3.8.5) using tmap-TDA (v1.2) and scikit-learn (v0.24.1). Directed acyclic graph (DAG) was created using DAGitty (v3.0) web interface tool (www.dagitty.net).

229 **References for supplementary material**

230 1. Dreborg S, Frew A. Position paper: Allergen standardization and skin tests. The European
231 Academy of Allergology and Clinical Immunology. Allergy. 1993;48(14 Suppl):48-82.
232 2. Mortensen MS, Brejnrod AD, Roggenbuck M, et al. The developing hypopharyngeal microbiota in
233 early life. Microbiome. 2016;4(1):70.
234 3. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010 [Available
235 from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
236 4. Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools
237 and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
238 5. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011.
239 2011;17(1):3.
240 6. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from
241 Illumina amplicon data. Nat Methods. 2016;13:581.
242 7. Wang Q, Garrity GM, Tiedje JM, et al. Naïve Bayesian Classifier for Rapid Assignment of rRNA
243 Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7.
244 8. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved
245 data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-D6.
246 9. Davis NM, Proctor DM, Holmes SP, et al. Simple statistical identification and removal of
247 contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):226.
248 10. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, et al. Sputum microbiome profiles identify severe
249 asthma phenotypes of relative stability at 12 to 18 months. J Allergy Clin Immunol. 2021;147(1):123-34.
250 11. Ward JH. Hierarchical Grouping to Optimize an Objective Function. J Am Stat Assoc.
251 1963;58(301):236-44.

12. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65.

13. Hartigan JA. Clustering Algorithms: John Wiley \& Sons, Inc.; 1975. 351 p.

14. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat 1974;3(1):1-27.

15. Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin. 1976;83(6):1072-80.

16. Kaufman L, Leonard Kaufman PJR, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis: Wiley; 1990.

17. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. J Am Stat Assoc. 1971;66(336):846-50.

18. Liao T, Wei Y, Luo M, et al. tmap: an integrative framework based on topological data analysis for population-scale microbiome stratification and association studies. Genome Biol. 2019;20(1):293.

19. Hennig C. Cluster-wise assessment of cluster stability. Comput Stat Data Anal. 2007;52(1):258-71.

20. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.

21. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300.

22. Beasley TM, Schumacker RE. Multiple Regression Approach to Analyzing Contingency Tables: Post Hoc and Planned Comparison Procedures. The Journal of Experimental Education. 1995;64(1):79-93.

23. Zhang Z. Variable selection with stepwise and best subset approaches. Ann Transl Med. 2016;4(7):136.

24. Liberzon A, Birger C, Thorvaldsdottir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.

25. Peña EA, Slate EH. Global Validation of Linear Model Assumptions. J Am Stat Assoc. 2006;101(473):341.

279 **Supplementary Tables:**

280 **Table E1:** list of 21 Hallmark gene sets from the Molecular Signatures Database (MSigDB) investigated using the gene set variation analysis
281 (GSVA).

Hallmark Name	Process Category	Description	Number of genes	Gene names
PEROXISOME	cellular component	peroxisomes	104	ABCD3, ACOT8, ACOX1, ACSL1, ECH1, ECI2, EHHADH, GSTK1, HSD17B4, MLYCD, PEX11A, RETSAT, SLC27A2, PEX13, PEX14, SCP2, HSD3B7, GNPAT, ABCD2, SLC25A17, PEX2, ACAA1, HAO2, HSD17B11, CRAT, PEX11B, LONP2, IDH1, FIS1, PEX6, ABCB4, SOD1, ABCB1, ISOC1, YWHAH, EPHX2, ABCD1, HMGCL, ACSL5, ALDH9A1, DHCR24, ELOVL5, NUDT19, PRDX5, CTPS1, IDE, SLC23A2, PEX5, BCL10, NR1I2, TSPO, CNBP, MSH2, DHRS3, DIO1, SLC25A4, PRDX1, IDI1, HRAS, MVP, ABCC8, CLN6, CAT, ACSL4, IDH2, ABCC5, SOD2, SLC35B2, FDPS, ALB, FADS1, STS, SMARCC1, ITGB1BP1, SIAH1, SLC25A19, CDK7, RXRG, ALDH1A1, UGT2B17, CADM1, SERPINA6, CLN8, RDH11, CTBP1, HSD11B2, TTR, ERCC3, ATXN1, SULT2B1, CRABP2, CRABP1, TOP2A, SCGB1A1, ERCC1, DLG4, PABPC1, FABP6, ABCB9, CACNA1B, SEMA3C, VPS4B, CEL, ESR2
MYOGENESIS	development	muscle differentiation	200	ACTA1, TNNI2, MYL1, TNNC1, TNNC2, MYH3, MYLPF, TNNT3, TNNT2, CASQ2, ACTC1, MYOM1, MYL4, MYBPH, MYH7, MYH8, ACTN2, TNNI1, CRYAB, SGCG, HRC, TNNT1, DES, MYOZ1, RYR1, CSRP3, ADAM12, ATP2A1, CKM, SVIL, MYOM2, MYL6B, TPM2, MYL2, CKMT2, BIN1, MYH1, ENO3, FLII, FXYD1, TPM3, DMD, IGFBP3, CHRNB1, SPEG, FHL1, ACTN3, TCAP, MYLK, MYL3, PYGM, LDB3, COX6A2, FABP3, MYL7, ITGB5, CHRNA1, SSPN, COL3A1, KCNH1, GJA5, MYF6, MYH2, MAPK12, PGAM2, MYOG, MYH4, AEBP1, HBEGF, MEF2C, NOS1, CNN3, IGFBP7, CACNA1H, GSN, CACNG1, PPFIA4, MB, SPHK1, SCHIP1, MEF2D, SMTN, CDKN1A, GAA, TPD52L1, HSPB2, SGCA, BDKRB2, COX7A1, COL4A2, PLXNB2, CTF1, COL15A1, KCNH2, AGRN, MYO1C, SIRT2, SGCD, SORBS1, VIPR1, FGF2, FKBP1B, TEAD4, CASQ1, KLF5, PDLIM7, AK1, TAGLN, RIT1, MEF2A, ANKRD2, AKT2, LAMA2, DENND2B, IFRD1, NCAM1, SYNGR2, PICK1, COL6A3, CRAT, DMPK, MYH11, MYH9, IGF1, CKB, FST, GPX3, PVALB, PTP4A3, ITGB1, HSPB8, ACHE, CHRNG, PKIA, NAV2, HDAC5, ATP6AP1, CLU, ACSL1, COL1A1, MRAS, PDE4DIP, AGL, ERBB3, GABARAPL2, CAV3, BHLHE40, PPP1R3C, LARGE1, LPIN1, CDH13, NQO1, PRNP, PSEN2, DAPK2, LSP1, SLC6A8, REEP1, EIF4A2, SH2B1, MYBPC3, ITGA7, SPTAN1, SORBS3, PC, SCD, GNAO1, STC2, CFD, RB1, BAG1, ABLIM1, PTGIS, MAPRE3, OCEL1, SH3BGR, NOTCH1, TSC2, TGFB1, FDPS, CAMK2B, EPHB3, GADD45B, APP, APLNR, ITGB4, SOD3, FOXO4, EFS, PFKM, COL6A2, DTNA, KIFC3, SPARC, SPDEF, SLN, WWTR1, CD36, ADCY9, APOD
COAGULATION	immune	blood coagulation cascade	138	F2, PROC, C1S, MMP14, F10, PLG, C1R, SERPINE1, SERPING1, C2, F12, CFI, MMP2, C3, F9, CTSO, TMPRSS6, MMP9, PROZ, MMP1, CFB, BMP1, VWF, FGA, F11, THBD, FURIN, MMP11, F8, LGMN, HPN, CTSB, MBL2, F13B, TIMP1, CTSV, MMP8, A2M, MMP10, MASP2, FGG, C8A, PLAU, ADAM9, GSN, CAPN2, PLAT, FN1, CTSH, OLR1, SPARC, RGN, C8B, LAMP2, C9, PROS1, CLU, F2RL2, CD9, CFH, PREP, LTA4H, GP9, DUSP14, CTSK, GDA, CASP9,

				APOC1, CRIP2, KLK8, SERPINC1, ANXA1, ITIH1, PRSS23, FBN1, MST1, F3, KLF7, CTSE, ITGB3, CAPN5, CSRP1, KLKB1, MSRB2, TFPI2, MMP3, MEP1A, GP1BA, APOA1, SERPINA1, TF, PEF1, GNG12, SIRT2, COMP, HTRA1, MAFF, LRP1, PF4, CPQ, MMP15, RABIF, HRG, C8G, APOC2, SERPINB2, RAPGEF3, CPN1, ITGA2, RAC1, DPP4, WDR1, THBS1, FYN, PDGFB, SH2B2, PECAM1, MMP7, DUSP6, CPB2, S100A13, CFD, DCT, LEFTY2, ANG, C1QA, HMGCS2, ISCU, PLEK, APOC3, P2RY1, USP11, TIMP3, ACOX2, HNF4A, S100A1, GNB2, ARF4
COMPLEMENT	immune	complement cascade	200	C2, C1S, CFB, C1R, SERPINE1, MMP14, SERPING1, CTSL, F5, MMP13, F7, CTSS, LGMN, PLG, C1QA, CASP1, GZMA, ADAM9, CALM3, C1QC, TIMP1, DPP4, KLK1, KLKB1, CD59, CR2, MMP15, LAP3, SPOCK2, F10, CTSB, SERPINA1, CTSO, CD40LG, CBLB, PDP1, C4BPB, PLEK, GP9, PLAUR, C3, F2, CASP4, STX4, CTSC, USP15, CR1, DUSP6, SERPINB2, GPD2, CFH, FN1, CD36, CA2, PSMB9, APOBEC3G, FCN1, GZMK, PDGFB, CLU, CASP10, LRP1, CTSD, S100A9, WAS, BRPF3, PLAT, CDA, MT3, CASP7, PRSS36, PFN1, GZMB, RNF4, ZEB1, CASP5, IRF1, CPQ, CDK5R1, ATOX1, PIK3CA, TMPRSS6, CPM, RCE1, CALM1, DOCK9, KYNU, RASGRP1, USP14, LCP2, GP1BA, KIF2A, GNB4, LCK, OLR1, PREP, MSRB1, LTA4H, ZFPM2, LYN, ACTN2, SIRT6, APOC1, PRKCD, ITGAM, DGKH, LTF, MAFF, KCNIP2, PCLO, DOCK10, SH2B3, RABIF, SRC, HPCAL4, CD46, PRDM4, GNAI3, C9, PPP2CB, IRF2, FYN, JAK2, PLA2G4A, PRCP, USP8, RHOG, L3MBTL4, LAMP2, PIM1, CXCL1, F3, GNAI2, CASP9, XPNPEP1, PLSCR1, IRF7, CD55, HSPA5, GNB2, DYRK2, PLA2G7, S100A12, GRB2, PHEX, GNGT2, DOCK4, MMP12, KCNIP3, FDX1, TIMP2, MMP8, FCER1G, RBSN, ANXA5, CTSV, GCA, EHD1, PRSS3, COL4A2, CSRP1, PIK3R5, SERPINC1, ANG, APOBEC3F, GATA3, DUSP5, CASP3, USP16, CP, PSEN1, LIPA, PCSK9, DGKG, GNG2, ME1, GMFB, SCG3, PPP4C, CCL5, CTSH, F8, APOA4, IL6, AKAP10, ERAP2, VCPIP1, HSPA1A, RAF1, NOTCH4, ADRA2B, CEBPB, HNF4A, LGALS3, TNFAIP3, CDH13, ITIH1, TFPI2, PIK3CG, S100A13
INTERFERON_ALPHA_RESPONSE	immune	interferon alpha response	97	MX1, ISG15, OAS1, IFIT3, IFI44, IFI35, IRF7, RSAD2, IFI44L, IFITM1, IFI27, IRF9, OASL, EIF2AK2, IFIT2, CXCL10, TAP1, SP110, DDX60, UBE2L6, USP18, PSMB8, IFIH1, BST2, LGALS3BP, ADAR, ISG20, GBP2, IRF1, PLSCR1, PSMB9, HERC6, SAMD9, CMPK2, IFITM3, RTP4, STAT2, SAMD9L, LY6E, IFITM2, HELZ2, CXCL11, TRIM21, PARP14, TRIM26, PARP12, NMI, RNF31, HLA-C, CASP1, TRIM14, TDRD7, DHX58, PARP9, PNPT1, TRIM25, PSME1, WARS1, EPSTI1, UBA7, PSME2, B2M, TRIM5, C1S, LAP3, LAMP3, GBP4, NCOA7, TMEM140, CD74, GMPR, PSMA3, PROCR, IL7, IFI30, IRF2, CSF1, IL15, CNP, TENT5A, IL4R, CMTR1, CD47, LPAR6, MOV10, CASP8, TXNIP, SLC25A28, SELL, TRAFD1, BATF2, RIPK2, CCRL2, NUB1, OGFR, MVB12A, ELF1
INTERFERON_GAMMA_RESPONSE	immune	interferon gamma response	200	STAT1, ISG15, IFIT1, MX1, IFIT3, IFI35, IRF7, IFIT2, OAS2, TAP1, EIF2AK2, RSAD2, MX2, IRF1, OAS3, TNFSF10, IRF9, CXCL10, IFI44, BST2, XAF1, SP110, OASL, PSMB8, IFI44L, IFITM3, DDX60, LGALS3BP, GBP4, IRF8, PSMB9, PML, IFIH1, UBE2L6, IFI27, ADAR, LY6E, STAT2, CXCL9, IL10RA, PLA2G4A, TRIM21, USP18, PTGS2, EPSTI1, C1S, DDX58, IL15, NLRC5, NMI, IDO1, PSMB10, CXCL11, ITGB7, SAMHD1, HERC6, CMPK2, SAMD9L, RTP4, PTPN2, PARP14, TNFAIP2, IFITM2, PLSCR1, SOCS1, CASP1, ICAM1, WARS1, PSME1, ISG20, IRF2, TRIM14, FCGR1A, MARCHF1, SOCS3, JAK2, HLA-DMA, PARP12, TNFAIP6, TRIM26, VCAM1, CD274, CIITA, NAMPT, SELP, GPR18, FPR1, HELZ2, PSME2, SERPING1, CCL5, RNF31, SOD2, TRIM25, LAP3, PSMA3, RNF213, PELI1, CFB, CD86, TXNIP, HLA-DQA1, GCH1, PNP, CCL7, PTPN6, SPPL2A, IL4R, PNPT1, DHX58, BTG1, CASP8, IFI30, CCL2, FGL2, CASP7, SECTM1, IL15RA, CD40, TRAFD1, HLA-DRB1, GBP6, LCP2, HLA-G, MT2A, RIPK1, KLRK1, UPP1, PSMB2, TDRD7, HIF1A, EIF4E3, VAMP8, PFKP, CD38, ZBP1, BANK1, TOR1B, RBCK1, PDE4B, MVP, IL7, BPGM, CMTR1, AUTS2, B2M, RIPK2, CD69, MYD88, PSMA2, PIM1, NOD1, CFH, TAPBP, SLC25A28, PTPN1, TNFAIP3, SSPN, NUP93, MTHFD2, CDKN1A, IRF4, NFKB1, BATF2, HLA-B, LATS2, IRF5, SLAMF7, ISOC1, P2RY14, STAT3, NCOA3, HLA-A,

				IL6, GZMA, IFNAR2, CD74, RAPGEF6, CASP4, FAS, OGFR, ARL4A, SRI, LYSMD2, CSF2RB, ST3GAL5, C1R, CASP3, CMKLR1, NFKBIA, METTL7B, ST8SIA4, XCL1, IL2RB, VAMP5, IL18BP, ZNFX1, ARID5B, APOL6, STAT4
IL6_JAK_STAT3_SIGNALING	immune	IL6 STAT3 signaling during acute phase response	87	IL4R, IL6ST, STAT1, IL1R1, CSF2RB, SOCS3, STAT3, OSMR, IL2RG, IFNGR1, TYK2, IL13RA1, TLR2, IFNGR2, IL10RB, IL6, IL1R2, IL3RA, IFNAR1, TNFRSF1A, MYD88, ACVR1B, CSF3R, ITGB3, REG1A, CXCL1, A2M, CSF2RA, IL15RA, IRF9, PDGFC, HAX1, BAK1, EBI3, INHBE, CRLF2, TNFRSF1B, CD14, PTPN1, PTPN2, IL1B, CSF1, IL18R1, TNF, PF4, CXCL13, LTBR, FAS, IL17RA, CXCL10, IL9R, STAM2, TNFRSF12A, STAT2, HMOX1, LEPR, CBL, CD9, CXCL3, TGFB1, MAP3K8, ITGA4, CD38, JUN, SOCS1, ACVRL1, PIM1, TNFRSF21, PIK3R5, GRB2, IRF1, DNTT, CSF2, IL2RA, PTPN11, IL12RB1, CCR1, CNTFR, PLA2G2A, CXCL9, CD44, IL7, CXCL11, CCL7, LTB, IL17RB, CD36
INFLAMMATORY_RESPONSE	immune	inflammation	200	CXCL10, CCL2, CCL5, FPR1, CCL20, IL1A, CXCL8, CCL7, CCL22, CXCL11, CCR7, EDN1, CD40, CXCL9, IL6, IL1B, TLR2, IL1R1, CD69, ICAM1, CCRL2, AQP9, EREG, C3AR1, GNA15, CMKLR1, PTGER4, LIF, IL15, NAMPT, OPRK1, ITGB8, PTAFR, ADM, PLAUR, NFKB1, INHBA, OSM, TNFSF10, TNFSF15, IFNGR2, ADGRE1, IL12B, CSF1, CXCL6, TNFRSF9, LYN, ACVR2A, LDLR, BDKRB1, HRH1, F3, BST2, PTGIR, CD55, CALCRL, CSF3, GPR132, IL4R, NLRP3, IL15RA, ADORA2B, GCH1, OLR1, PTGER2, CSF3R, MYC, RELA, TNFAIP6, IL7R, IL18, GABBR1, CD82, TNFSF9, NMUR1, IL2RB, TLR1, LPAR1, IRAK2, RIPK2, MMP14, P2RX7, SLC11A2, SELL, P2RY2, ABCA1, FFAR2, PROK2, GNAI3, TACR1, SLC7A1, CDKN1A, CYBB, TIMP1, HBEGF, SCARF1, EBI3, NFKBIA, SRI, SLC7A2, CCL17, TLR3, APLNR, OSMR, IL10RA, PSEN1, GPR183, ATP2B1, TNFRSF1B, BEST1, GPC3, SCN1B, ACVR1B, HPN, SEMA4D, KLF6, CD48, CXCR6, SLC1A2, GP1BA, TAPBP, RGS16, SLAMF1, LCK, HIF1A, AHR, NMI, RHOG, TPBG, NPFFR2, IFNAR1, ICOSLG, RASGRP1, IFITM1, KCNJ2, LY6E, IL18R1, IL10, KCNA3, HAS2, DCBLD2, LAMP3, VIP, CD70, RGS1, SLC31A1, ADRM1, KCNMB2, SERPINE1, MXD1, AXL, MEFV, PVR, CCL24, PDE4B, LCP2, PDPN, IRF7, MET, ATP2A2, SLC31A2, FZD5, ITGA5, SGMS2, MARCO, CD14, EIF2AK2, ROS1, ATP2C1, NDP, BTG2, MSR1, PTPRE, RNF144B, PCDH7, SPHK1, IL18RAP, RTP4, RAF1, CHST2, ITGB3, KIF1B, SELE, NOD2, C5AR1, EMP3, CLEC5A, TACR3, SLC4A4, MEP1A, SELENOS, LTA, PIK3R5, STAB1, IRF1, ICAM4, P2RX4, ABI1, CX3CL1, SLC28A2
OXIDATIVE_PHOSPHORYLATION	metabolic	oxidative phosphorylation and citric acid cycle	200	NDUFS3, UQCRB, NDUFS2, SDHA, UQCRC1, NDUFA9, NDUFS4, NDUFS1, NDUFA2, NDUFS8, SDHB, NNT, ATP5PO, ATP5MC3, NDUFS7, ATP5F1A, NDUFV1, COX5B, UQCRH, NDUFA1, ATP5F1C, ATP5F1B, COX7B, SDHD, CYCS, NDUFA6, NDUFAB1, COX8A, ACO2, ATP5MC1, CYC1, NDUFB6, ATP5F1E, COX5A, UQCRC2, COX6A1, ATP5F1D, COX6C, ATP5PF, NDUFB3, IDH3B, OGDH, NDUFB8, SURF1, COX6B1, NDUFB5, NDUFA4, NDUFB1, COX4I1, COX7C, UQCRFS1, SDHC, ATP6V1F, COX7A2, SUCLG1, NDUFS6, NDUFA7, FH, NDUFV2, OXA1L, NDUFC1, UQCR11, NDUFA5, CS, ATP6V1G1, ATP5PB, HCCS, HADHB, ATP5PD, PDHA1, NDUFA8, DLD, OPA1, SLC25A11, ATP5ME, PDHB, ATP5MF, NDUFB7, IDH2, MTX2, VDAC3, MDH1, ATP5MC2, IMMT, MDH2, SLC25A3, ATP6V1D, VDAC2, ACADM, COX7A2L, TIMM17A, ATP6V1E1, NDUFA3, SLC25A6, IDH3G, ACADVL, ETFA, TIMM9, IDH3A, TIMM8B, ATP6AP1, TIMM13, UQCRQ, ABCB7, VDAC1, ATP5MG, PHB2, DECR1, SUCLA2, GOT2, DLAT, ATP6V1H, NDUFB2, FDX1, HADHA, ATP6V1C1, MAOB, NDUFB4, UQCR10, ETFDH, GPX4, PDHX, MFN2, AIFM1, ACAA2, ETFB, COX11, ECHS1, PMPCA, ATP6V0B, SLC25A5, DLST, COX15, CYB5A, ALAS1, SLC25A4, CPT1A, SLC25A20, MTRR, COX17, CYB5R3, TOMM22, ACAT1, MRPS11, ATP6V0C, PDK4, TIMM10, LDHA, ECI1, MRPL11, FXN, MRPS12, COX10, RHOT1, ACAA1, ACADSB, LDHB, MRPS30, ATP1B1, BDH2, SLC25A12, TIMM50, MRPL34, ISCA1, MRPL35, IDH1, HSPA9, MRPL15, MRPS15, TOMM70, TCIRG1, ISCU, POLR2F, NQO2, NDUFC2, MRPS22, POR, ATP6V0E1, PHYH, MPC1, GPI, AFG3L2, HSD17B10, CASP7, PRDX3, MGST3, HTRA2, BCKDHA, LRPPRC, RETSAT, ECH1, RHOT2, BAX, MTRF1, GLUD1, SUPV3L1, GRPEL1, PDP1, ALDH6A1, OAT

HYPOXIA	path way	response to hypoxia; HIF1A targets	200	PGK1, PDK1, GBE1, PFKL, ALDOA, ENO2, PGM1, NDRG1, HK2, ALDOC, GPI, MXI1, SLC2A1, P4HA1, ADM, P4HA2, ENO1, PFKP, AK4, FAM162A, PFKFB3, VEGFA, BNIP3L, TPI1, ERO1A, KDM3A, CCNG2, LDHA, GYS1, GAPDH, BHLHE40, ANGPTL4, JUN, SERPINE1, LOX, GCK, PPFIA4, MAFF, DDIT4, SLC2A3, IGFBP3, NFIL3, FOS, RBPJ, HK1, CITED2, ISG20, GALK1, WSB1, PYGM, STC1, ZNF292, BTG1, PLIN2, CSRP2, VLDLR, JMJD6, EXT1, F3, PDK3, ANKZF1, UGP2, ALDOB, STC2, ERRFI1, ENO3, PNRC1, HMOX1, PGF, GAPDHS, CHST2, TMEM45A, BCAN, ATF3, CAV1, AMPD3, GPC3, NDST1, IRS2, SAP30, GAA, SDC4, STBD1, IER3, PKLR, IGFBP1, PLAUR, CAVIN3, CCN5, LARGE1, NOCT, S100A4, RRAGD, ZFP36, EGFR, EDN2, IDS, CDKN1A, RORA, DUSP1, MIF, PPP1R3C, DPYSL4, KDELR3, DTNA, ADORA2B, HS3ST1, CAVIN1, NR3C1, KLF6, GPC4, CCN1, TNFAIP3, CA12, HEXA, BGN, PPP1R15A, PGM2, PIM1, PRDX5, NAGK, CDKN1B, BRS3, TKTL1, MT1E, ATP7A, MT2A, SDC3, TIPARP, PKP1, ANXA2, PGAM2, DDIT3, PRKCA, SLC37A4, CXCR4, EFNA3, CP, KLF7, CCN2, CHST3, TPD52, LXN, B4GALNT2, PPARGC1A, BCL2, GCNT2, HAS1, KLHL24, SCARB1, SLC25A1, SDC2, CASP6, VHL, FOXO3, PDGFB, B3GALT6, SLC2A5, SRPX, EFNA1, GLRX, ACKR3, PAM, TGFBI, DCN, SIAH2, PLAC8, FBP1, TPST2, PHKG1, MYH9, CDKN1C, GRHPR, PCK1, INHA, HSPA5, NDST2, NEDD4L, TPBG, XPNPEP1, IL6, SLC6A6, MAP3K1, LDHC, AKAP12, TES, KIF5A, LALBA, COL5A1, GPC1, HDLBP, ILVBL, NCAN, TGM2, ETS1, HOXB9, SELENBP1, FOSL2, SULT2B1, TGFB3
REACTIVE_ OXYGEN_S PECIES_PA THWAY	path way	reactive oxygen species pathway	49	GSR, PRDX2, TXNRD1, SOD1, GCLC, CAT, GPX4, MPO, PRDX1, PRDX6, GCLM, TXN, SOD2, PRDX4, SELENOS, PDLIM1, TXNRD2, LAMTOR5, G6PD, MSRA, MBP, SRXN1, NQO1, OXSR1, GLRX2, HMOX2, FES, PFKP, NDUFB4, GPX3, SCAF4, CDKN2D, MGST1, ABCC1, FTL, ATOX1, STK25, EGLN2, ERCC2, SBNO2, JUNB, PTPA, LSP1, NDUFA6, HHEX, GLRX, NDUFS2, PRNP, IPCEF1
P53_PATH WAY	prolif erati on	p53 pathway	200	CDKN1A, BTG2, MDM2, CCNG1, FAS, TOB1, GADD45A, PHLDA3, TAP1, CDKN2B, SFN, ZMAT3, DDB2, EI24, PERP, DDIT4, ATF3, BAX, SESN1, FDXR, PIDD1, SAT1, CDKN2A, AEN, PPM1D, FOXO3, BTG1, TXNIP, SLC19A2, IER3, TP53, NOTCH1, RRAD, DCXR, NINJ1, FOS, S100A10, FBXW7, PLK3, XPC, AK1, TRIAP1, BAK1, CCND2, APAF1, PDGFA, NDRG1, RALGDS, SERPINB5, DGKA, CYFIP2, KLF4, IP6K2, ADA, TNFSF9, KLK8, PRKAB1, STOM, SOCS1, PLK2, CASP1, RCHY1, PCNA, RPS27L, RB1, PPP1R15A, POLH, RAB40C, ERCC5, CDH13, DDIT3, TGFB1, FUCA1, H2AW, TP63, HSPA4L, EPHX1, ABHD4, H2AJ, TRAF4, CTSF, IER5, TAX1BP3, TSC22D1, RGS16, HRAS, PVT1, ZFP36L1, SDC1, WRAP73, RAP2B, COQ8A, PLXNB2, HEXIM1, PROCR, SPHK1, NOL8, DEF6, SP1, JAG2, CCP110, DRAM1, RPL18, CEBPA, KIF13B, MXD1, TRAFD1, OSGIN1, ANKRA2, HBEGF, RAD51C, ITGB4, TRIB3, INHBB, VAMP8, VWA5A, ABAT, CTSD, PTPRE, HINT1, ALOX15B, VDR, ELP1, WWP1, NUPR1, TCN2, CDKN2AIP, POM121, ZNF365, EPHA2, JUN, MAPKAPK3, SLC7A11, SERTAD3, IRAK1, ABCC5, UPP1, CD81, LIF, RPL36, FAM162A, RNF19B, CDK5R1, STEAP3, PITPNC1, SLC35D1, TPRKB, RRP8, IL1A, PMM1, TPD52L1, IFI30, GM2A, RPS12, EPS8L2, TM4SF1, BLCAP, MKNK2, CD82, RACK1, ST14, TCHH, HDAC3, CCNK, RHBDF2, NUDT15, TGFA, ISCU, LDHB, TNNI1, FGF13, RXRA, APP, F2R, CLCA2, CGRRF1, CSRNP2, ACVR1B, BMP2, GPX2, KRT17, TM7SF3, ZBTB16, PTPN14, TSPYL2, SLC3A2, S100A4, GLS2, CCND3, PRMT2, NHLH2, BAIAP2, RETSAT, IRAG2, H1-2, DNTTIP2, MXD4, SEC61A1, HMOX1, RAD9A
IL2_STAT5 _SIGNALIN G	signa ling	IL2 STAT5 signaling	199	SOCS2, CISH, PIM1, IL2RA, TNFRSF4, SOCS1, TNFRSF9, XBP1, RRAGD, HK2, PHLDA1, IL2RB, CTLA4, NFIL3, CD83, IKZF2, IL10, TNFRSF18, DHRS3, ECM1, ADAM19, SLC2A3, HIPK2, BATF3, BHLHE40, PTGER2, DENND5A, ITIH5, PHTF2, GADD45B, NRP1, NCOA3, CD79B, AHR, TNFRSF1B, NDRG1, BCL2L1, GABARAPL1, LIF, TIAM1, BMPR2, MAP3K8, RHOB, MYC, S100A1, ETFBKMT, CAPG, ST3GAL4, PENK, IRF4, CST7, WLS, TLR7, IKZF4, GBP4, RGS16, SPP1, IL13, SLC29A2, NFKBIZ, IL4R, MXD1, CSF2, FAH, CTSZ, ITGAE, MUC1, MAPKAPK2, TNFRSF21, NT5E,

				FLT3LG, CCND2, TRAF1, LCLAT1, IL3RA, CYFIP1, BCL2, FGL2, PRNP, EEF1AKMT1, PUS1, ITGAV, NCS1, DCPS, AMACR, FAM126B, PTH1R, ODC1, IGF1R, PTCH1, ENO3, CD81, MAFF, EMP1, CDKN1C, CAPN3, IL1R2, SYT11, TTC39B, ANXA4, BATF, P4HA1, GPR65, SLC1A5, IGF2R, CKAP4, CCR4, CD44, P2RX4, GATA1, KLF6, ARL4A, HOPX, GPR83, ITGA6, CD48, DRC1, SELP, GLIPR2, SMPDL3A, PLSCR1, FURIN, SERPINB6, TNFSF11, GPX4, LRRC8C, CCNE1, CASP3, SH3BGRL2, SNX9, PLEC, BMP2, ICOS, ALCAM, LTB, ENPP1, IL1RL1, MYO1C, IFNGR1, PLIN2, IL18R1, AHNAK, PRKCH, TNFRSF8, SYNGR2, GALM, POU2F1, EOMES, NOP2, PTRH2, RHOH, CDC6, MYO1E, CXCL10, SNX14, IRF6, IL10RA, MAP6, TNFSF10, SPRED2, SELL, SERPINC1, CDCP1, RORA, COCH, CSF1, F2RL2, UCK2, CA2, IFITM3, UMPS, HUWE1, COL6A1, ABCB1, RNH1, IRF8, GUCY1B1, AHCY, PRAF2, GSTO1, TWSG1, CDC42SE2, PLAGL1, APLP1, PLPP1, SPRY4, SCN9A, SHE, PDCD2L, CCND3, LRIG1, SWAP70, SLC39A8, RABGAP1L, TGM2, PNP, AGER, ETV4, CD86
KRAS_SIGNALING_UP	signaling	KRAS signaling, upregulated genes	200	ANGPTL4, ITGA2, SPRY2, HBEGF, RBP4, HSD11B1, ETV4, GLRX, DUSP6, SCG5, ETV5, ITGB2, AKT2, PPBP, G0S2, GABRA3, IRF8, BIRC3, FGF9, DCBLD2, INHBA, TFPI, TSPAN1, ADAM8, SLPI, PRKG2, MMP11, MMP10, TMEM158, TNFAIP3, PRDM1, GALNT3, ETS1, MMP9, WNT7A, IGFBP3, SPP1, ETV1, CLEC4A, CCND2, TSPAN7, ITGBL1, EMP1, CDADC1, KIF5C, TRIB2, SDCCAG8, PCP4, CFHR2, ALDH1A2, NR0B2, ALDH1A3, AMMECR1, SATB1, GUCY1A1, CSF2, APOD, TOR1AIP2, CMKLR1, TMEM176B, ADGRA2, LAPTM5, CD37, CAB39L, CIDEA, ZNF639, IL1B, GYPC, LY96, FLT4, SPON1, BMP2, PLEK2, IGF2, NR1H4, SNAP25, ACE, PRRX1, C3AR1, TRAF1, TLR8, ID2, TMEM100, PLAUR, GADD45G, CBX8, SCN1B, PTBP2, NAP1L2, AKAP12, PLAT, SCG3, ANO1, IL1RL2, CXCL10, ATG10, YRDC, HDAC9, PEG3, SEMA3B, TNNT2, LIF, CFB, BTC, PPP1R15A, PTPRR, CCL20, ARG1, RETN, KLF4, MMD, PDCD1LG2, H2BC3, HOXD11, TRIB1, F2RL1, ANXA10, TSPAN13, MTMR10, CFH, LAT2, ERO1A, RELN, KCNN4, TMEM176A, MAP4K1, PTGS2, IL33, MAFB, LCP1, NGF, CA2, SERPINA3, RGS16, CTSS, USP12, CPE, SPARCL1, ABCB1, USH1C, CSF2RA, BTBD3, IL2RG, DNMBP, IL10RA, EREG, PRELID3B, EPHB2, FBXO4, CROT, MPZL2, ANKH, CBR4, DOCK2, GPRC5B, RABGAP1L, MALL, STRN, ST6GAL1, PIGR, VWA5A, PSMB8, F13A1, NRP1, SOX9, JUP, ADGRL4, ZNF277, EPB41L3, PCSK1N, FUCA1, PLVAP, ADAM17, AVL9, ADAMDEC1, HKDC1, MAP7, IL7R, RBM4, BPGM, ENG, GFPT2, PLAU, GNG11, PTCD2, MAP3K1, CBL, CXCR4, NIN, IKZF1, WDR33, MYCN, FCER1G, PECAM1, CCSER2, SNAP91, EVI5, TNFRSF1B, GPNMB, TPH1
MTORC1_SIGNALING	signaling	mTORC1 signaling	200	FADS1, DDIT4, CALR, HK2, PGK1, SLC7A5, CTSC, ACSL3, SLC1A5, M6PR, TFRC, DDIT3, TMEM97, IFRD1, PLOD2, TUBA4A, PSAT1, CORO1A, LDHA, MTHFD2, FADS2, VLDLR, WARS1, SCD, P4HA1, ACTR2, IDH1, SLC2A1, GBE1, SERPINH1, NUPR1, PSMG1, PSPH, NAMPT, CDKN1A, BHLHE40, HSPA9, HSPA5, EGLN3, LGMN, PNP, XBP1, SLA, DDX39A, HSPE1, ACLY, SLC7A11, SSR1, GLA, SQSTM1, PDK1, PSMC2, PRDX1, SERP1, TRIB3, NFIL3, HMGCS1, GOT1, TPI1, ELOVL6, ASNS, PSMD14, PSMA4, PPA1, HPRT1, AURKA, HMGCR, GAPDH, DHFR, DHCR7, IMMT, UCHL5, YKT6, INSIG1, SQLE, IGFBP5, IFI30, CYP51A1, FGL2, ENO1, IDI1, CYB5B, SHMT2, TXNRD1, G6PD, SLC9A3R1, RAB1A, EBP, PNO1, PIK3R3, ACTR3, LDLR, SLC2A3, UBE2D3, ELOVL5, CACYBP, EDEM1, ATP6V1D, TES, TM7SF2, PSMA3, ITGB2, AK4, SLC1A4, TOMM40, SLC6A6, PPIA, ADD3, ME1, CCNF, SLC37A4, ALDOA, BTG2, UFM1, CCNG1, STC1, NMT1, PSMC6, FDXR, RRM2, DHCR24, PSMC4, CTH, PSME3, CFP, POLR3G, ACACA, QDPR, MCM2, PSMD12, SC5D, CANX, RPN1, HSPA4, NIBAN1, TBK1, SEC11A, BCAT1, PSMB5, PSMD13, PGM1, PLK1, GLRX, COPS5, ETF1, GSK3B, NUP205, SORD, PHGDH, GMPS, RRP9, EEF1E1, LTA4H, SDF2L1, FKBP2, RDH11, CXCR4, MLLT11, GCLC, TCEA1, MAP2K3, HSPD1, SYTL2, MCM4, PPP1R15A, USO1, NFKBIB, UNG, GTF2H1, RPA1, HSP90B1, ERO1A, GSR, PITPNB, EPRS1, SRD5A1, TUBG1, MTHFD2L, ADIPOR2, NUFIP1, CDC25A,

				PDAP1, STARD4, BUB1, ARPC5L, GPI, EIF2S2, CD9, ATP2A2, GGA2, HMBS, RIT1, SKAP2, STIP1, DAPP1, ABCF2, NFYC, ATP5MC1, PFKL, CCT6A
NOTCH_SIGNALING	signaling	Notch signaling	32	JAG1, NOTCH3, NOTCH2, APH1A, HES1, CCND1, FZD1, PSEN2, FZD7, DTX1, DLL1, FZD5, MAML2, NOTCH1, PSENEN, WNT5A, CUL1, WNT2, DTX4, SAP30, PPARD, KAT2A, HEYL, SKP1, RBX1, TCF7L2, ARRB1, LFNG, PRKCA, DTX2, ST3GAL6, FBXW11
PI3K_AKT_MTOR_SIGNALING	signaling	PI3K signaling via AKT to mTORC1	105	MAPK8, PIK3R3, GRB2, NFKBIB, MAP2K6, MAPK9, AKT1, MAPK1, PLCG1, TRIB3, GSK3B, MAP2K3, CDKN1A, RAC1, RIPK1, AKT1S1, ACTR2, PRKAR2A, YWHAB, HRAS, PDK1, PIKFYVE, TBK1, ACTR3, E2F1, MYD88, ITPR2, SQSTM1, RPS6KA1, PTPN11, MAPKAP1, PLCB1, RAF1, CAMK4, RPTOR, CFL1, CDK4, TRAF2, GNGT1, UBE2N, ADCY2, CDKN1B, VAV3, FGF6, ECSIT, RALB, ARF1, MKNK1, CDK1, PTEN, ARHGDIA, GRK2, FGF17, DDIT3, IRAK4, TIAM1, CDK2, SFN, PRKCB, GNA14, EIF4E, CLTC, TSC2, FGF22, PPP1CA, DUSP3, HSP90B1, IL4, STAT2, SLA, EGFR, PLA2G12A, MAPK10, CALR, THEM4, RIT1, MKNK2, PPP2R1B, CAB39L, ARPC3, PITX2, NCK1, IL2RG, PFN1, FASLG, NOD1, DAPP1, UBE2D3, CAB39, AP2M1, MAP3K7, PRKAG1, CSNK2B, PRKAA2, ATF1, SLC2A1, PIN1, TNFRSF1A, LCK, RPS6KA3, NGF, CXCR4, ACACA, SMAD2, PAK4
HEDGEHOG_SIGNALING	signaling	Hedgehog signaling	36	SHH, PTCH1, NRCAM, NRP1, SCG2, AMOT, UNC5C, ADGRG1, HEY1, GLI1, THY1, SLIT1, CDK6, HEY2, NRP2, TLE3, TLE1, L1CAM, PLG, NKX6-1, NF1, RASA1, ETS2, RTN1, CRMP1, MYH9, VEGFA, CELSR1, CNTFR, ACHE, PML, CDK5R1, VLDLR, OPHN1, LDB1, DPYSL2
TGF_BETA_SIGNALING	signaling	TGF beta signaling	54	TGFBR1, SMAD7, TGFB1, SMURF2, SMURF1, BMPR2, SKIL, SKI, ACVR1, PMEPA1, NCOR2, SERPINE1, JUNB, SMAD1, SMAD6, PPP1R15A, TGIF1, FURIN, SMAD3, FKBP1A, MAP3K7, BMPR1A, CTNNB1, HIPK2, KLF10, BMP2, ENG, APC, PPM1A, XIAP, CDH1, ID1, LEFTY2, CDKN1C, TRIM33, RAB31, TJP1, SLC20A1, CDK9, ID3, NOG, ARID4B, IFNGR2, ID2, PPP1CA, SPTBN1, WWTR1, BCAR3, THBS1, FNTA, HDAC1, UBE2D3, LTBP2, RHOA
TNFA_SIGNALING_VIA_NFKB	signaling	TNFA signaling via NFkB	200	JUNB, CXCL2, ATF3, NFKBIA, TNFAIP3, PTGS2, CXCL1, IER3, CD83, CCL20, CXCL3, MAFF, NFKB2, TNFAIP2, HBEGF, KLF6, BIRC3, PLAUR, ZFP36, ICAM1, JUN, EGR3, IL1B, BCL2A1, PPP1R15A, ZC3H12A, SOD2, NR4A2, IL1A, RELB, TRAF1, BTG2, DUSP1, MAP3K8, ETS2, F3, SDC4, EGR1, IL6, TNF, KDM6B, NFKB1, LIF, PTX3, FOSL1, NR4A1, JAG1, CCL4, GCH1, CCL2, RCAN1, DUSP2, EHD1, IER2, REL, CFLAR, RIPK2, NFKBIE, NR4A3, PHLDA1, IER5, TNFSF9, GEM, GADD45A, CXCL10, PLK2, BHLHE40, EGR2, SOCS3, SLC2A6, PTGER4, DUSP5, SERPINB2, NFIL3, SERPINE1, TRIB1, TIPARP, RELA, BIRC2, CXCL6, LITAF, TNFAIP6, CD44, INHBA, PLAU, MYC, TNFRSF9, SGK1, TNIP1, NAMPT, FOSL2, PNRC1, ID2, CD69, IL7R, EFNA1, PHLDA2, PFKFB3, CCL5, YRDC, IFNGR2, SQSTM1, BTG3, GADD45B, KYNU, G0S2, BTG1, MCL1, VEGFA, MAP2K3, CDKN1A, CCN1, TANK, IFIT2, IL18, TUBB2A, IRF1, FOS, OLR1, RHOB, AREG, NINJ1, ZBTB10, PLPP3, KLF4, CXCL11, SAT1, CSF1, GPR183, PMEPA1, PTPRE, TLR2, ACKR3, KLF10, MARCKS, LAMB3, CEBPB, TRIP10, F2RL1, KLF9, LDLR, TGIF1, RNF19B, DRAM1, B4GALT1, DNAJB4, CSF2, PDE4B, SNN, PLEK, STAT5A, DENND5A, CCND1, DDX58, SPHK1, CD80, TNFAIP8, CCNL1, FUT4, CCRL2, SPSB1, TSC22D1, B4GALT5, SIK1, CLCF1, NFE2L2, FOSB, PER1, NFAT5, ATP2B1, IL12B, IL6ST, SLC16A6, ABCA1, HES1, BCL6, IRS2, SLC2A3, CEBPD, IL23A, SMAD3, TAP1, MSC, IFIH1, IL15RA, TNIP2, BCL3, PANX1, FJX1, EDN1, EIF1, BMP2, DUSP4, PDLIM5, ICOSLG, GFPT2, KLF2, TNC, SERPINB8, MXD1
WNT_BETA_CATENIN_SIGNALING	signaling	canonical beta catenin pathway	42	MYC, CTNNB1, JAG2, NOTCH1, DLL1, AXIN2, PSEN2, FZD1, NOTCH4, LEF1, AXIN1, NKD1, WNT5B, CUL1, JAG1, MAML1, KAT2A, GNAI1, WNT6, PTCH1, NCOR2, DKK4, HDAC2, DKK1, TCF7, WNT1, NUMB, ADAM17, DVL2, PPARD, NCSTN, HDAC5, CCND2, FRAT1, CSNK1E, RBPJ, FZD8, TP53, SKP2, HEY2, HEY1, HDAC11

282 This table is part of the Hallmark gene sets table from Molecular Signatures Database (MSigDB) release version 7.4 as previously described in [24] (Liberzon *et*
283 *al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-425).

284 **Table E2:** A stepwise multiple logistic regression model of having atopic dermatitis diagnosis.

Atopic dermatitis diagnosis							
Characteristic#	**n†**	**OR**	**SE**	**95% CI**	**Overall *P*-value‡**	***P*-value**	***q*-value**
Taxa-driven clusters¶					**8.76 x 10^{-3}**		
• **Cluster 2**	**87/239**	**3.89**	**0.411**	**1.76, 8.85**		**9.42 x 10^{-4}**	**0.008**
• Cluster 3	21/239	1.79	0.571	0.59, 5.65		0.308	0.308
• Cluster 4	81/239	2.18	0.394	1.01, 4.78		0.047	0.107
Ethnicity (White Caucasian)	192/239	0.46	0.420	0.19, 1.02		0.068	0.123
Season of inclusion*					0.063		
• Spring	53/239	0.53	0.441	0.22, 1.25		0.146	0.187
• Summer	60/239	0.57	0.420	0.25, 1.29		0.178	0.200
• Winter	54/239	0.32	0.433	0.13, 0.74		0.009	0.026
Library size	239	1.00	0.000	1.00, 1.00		0.102	0.153
Plate ID (Plate 2)	84/239	2.60	0.348	1.34, 5.26		0.006	0.026
	McFadden's R^2 = 0.103 , Nagelkerke R^2 = 0.170, AIC = 292.2						

285 AIC: Akaike information criterion, CI: confidence interval, SE: standard error, and OR: odds ratio. † n=2/241 patients
286 with missing atopic dermatitis diagnosis were excluded from the regression analysis. # Variables in the model were
287 selected according to a stepwise process. ¶ Cluster 1 (n=50/239, with the lowest % of atopy diagnosis) was the
288 reference group for comparison in the stepwise process. * Fall season (n=72/239) was randomly selected as a
289 reference group for comparison in the stepwise process. ‡ The overall *P*-values were computed for variables with
290 >2 categories, in which all the variable's categories were taken into account (without a reference group selection).
291 The microbiota clusters with the statistically significant q-values after adjusting for covariates are highlighted in
292 bold-face. *q*-value < 0.05 is considered statistically significant.

293 **Table E3:** A stepwise multiple linear regression model of FEV1 post-salbutamol % predicted.

FEV_1 % predicted post-salbutamol							
Characteristic#	**n†**	**β**	**SE**	**95% CI**	**Overall *P*-value‡**	***P*-value**	***q*-value**
Taxa-driven clusters¶					**0.030**		
• Cluster 1	29/143	6.3	3.76	-1.1, 14		0.094	0.148
• **Cluster 3**	**10/143**	**15**	**5.51**	**4.3, 26**		**0.007**	**0.037**
• Cluster 4	45/143	1.6	3.17	-4.7, 7.8		0.622	0.665
BMI z-score	143	2.1	1.10	-0.07, 4.3		0.058	0.118
Centre of Inclusion*					0.028		
• Centre B	11/143	2.5	5.66	-8.7, 14		0.665	0.665
• Centre D	4/143	-16	8.45	-33, 0.78		0.061	0.118
• Centre E	35/143	-11	3.89	-19, -3.3		0.005	0.037
• Centre F	30/143	-7.7	4.15	-16, 0.47		0.064	0.118
• Centre G	24/143	-3.9	4.41	-13, 4.9		0.383	0.469
OCS	27/143	-5.5	3.70	-13, 1.8		0.137	0.188
NCS	17/143	10	4.22	1.8, 19		0.017	0.064
R^2 = 0.180; adjusted R^2 = 0.112, AIC = 1209.8, overall *P*-value = 0.005							

294 β: beta regression co-efficient, AIC: Akaike information criterion, BMI: body mass index, CI: confidence interval,
295 NCS: nasal corticosteroids, OCS: oral corticosteroids, and SE: standard error. † Only older children with available
296 spirometry data (n=143/241) were included in the regression analysis. # Variables in the model were selected
297 according to a stepwise process. ¶ Cluster 2 (n=59/143, with the lowest FEV_1 post-salbutamol % predicted median
298 value) was the reference group for comparison in the stepwise process. *Inclusion center C (n=39/143) was
299 randomly selected as a reference group in the stepwise process. ‡ The overall *P*-values were computed for
300 variables with >2 categories, in which all the variable's categories were taken into account (without a reference
301 group selection). The microbiota clusters with the statistically significant q-values after adjusting for covariates are
302 highlighted in bold-face.

303

304 **Table E4:** A stepwise multiple logistic regression model of grass pollen sensitization.

Grass pollen sensitization							
Characteristic#	**n†**	**OR**	**SE**	**95% CI**	**Overall *P*-value‡**	***P*-value**	***q*-value**
Taxa-driven clusters¶					0.101		
• **Cluster 2**	**81/219**	**3.14**	**0.470**	**1.27, 8.10**		**0.015**	**0.038**
• Cluster 3	18/219	2.07	0.691	0.53, 8.15		0.292	0.292
• Cluster 4	74/219	2.07	0.481	0.82, 5.46		0.131	0.176
Age	219	1.28	0.038	1.19, 1.38		7.6×10^{-11}	<0.001
Sex (female)	86/219	0.60	0.352	0.30, 1.19		0.147	0.176
Ethnicity (White Caucasian)	176/219	0.39	0.398	0.18, 0.85		0.019	0.038
McFadden's R^2 = 0.233, Nagelkerke R^2 = 0.366, AIC = 243.9							

305 AIC: Akaike information criterion, CI: confidence interval, SE: standard error, and OR: odds ratio. † only participants
306 with grass pollen atopy sensitization (n=219/241), by either a skin prick test and/or an allergen-specific IgE, were
307 included in the regression analysis. # Variables in the model were selected according to a stepwise process. ¶
308 Cluster 1 (n=46/219, with the lowest % of grass atopy diagnosis) was the reference group for comparison in the
309 stepwise process. ‡ The overall *P*-value was computed for taxa-driven clusters, in which all clusters were taken into
310 account (without a reference group selection). The microbiota clusters with the statistically significant q-values
311 after adjusting for covariates are highlighted in bold-face. *q*-value < 0.05 is considered statistically significant.

312 **Table E5:** Stepwise multiple linear regression models of the statistically significant hallmark pathways.

Transcriptomic pathways							
Wnt/β-Catenin signaling							
Characteristic#	**n†**	**β**	**SE**	**95% CI**	**Overall *P*-value‡**	***P*-value**	***q*-value**
Taxa-driven clusters¶					**4.12 x 10^{-05}**		
• Cluster 2	69/188	0.08	0.033	0.01, 0.14		0.024	0.112
• Cluster 3	15/188	0.07	0.050	-0.03, 0.17		0.168	0.235
• **Cluster 4**	**64/188**	**0.16**	**0.034**	**0.10, 0.23**		**2.95 x 10^{-06}**	**<0.001**
Age	188	0.01	0.003	0.01, 0.02		3.41 x 10^{-05}	<0.001
Ethnicity (White Caucasian)	150/188	0.06	0.031	-0.01, 0.12		0.072	0.203
BMI z-score	188	0.02	0.011	0.00, 0.04		0.070	0.203
Centre of Inclusion*					0.070		
• Centre A	16/188	0.08	0.052	-0.02, 0.18		0.129	0.226
• Centre E	50/188	0.02	0.038	-0.06, 0.09		0.643	0.643
• Centre B	17/188	0.04	0.050	-0.06, 0.13		0.477	0.513
• Centre D	3/188	-0.13	0.099	-0.32, 0.07		0.197	0.251
• Centre F	33/188	-0.05	0.041	-0.13, 0.03		0.259	0.302
• Centre G	32/188	-0.06	0.040	-0.14, 0.01		0.111	0.226
NCS	12/188	-0.08	0.050	-0.18, 0.02		0.119	0.226
Plate ID (Plate 2)	69/188	-0.04	0.027	-0.09, 0.01		0.151	0.235
R^2 = 0.245; adjusted R^2 = 0.184, AIC = -137.5, overall *P*-value = 5.5 x 10^{-6}							
TGF-β *signaling*							
Characteristic#	**n†**	**β**	**SE**	**95% CI**	**Overall *P*-value‡**	**p-value**	**q-value**
Taxa-driven clusters¶					**0.002**		
• **Cluster 2**	**69/188**	**0.13**	**0.032**	**0.06, 0.19**		**1 x 10^{-4}**	**<0.001**
• Cluster 3	15/188	0.08	0.049	-0.02, 0.18		0.101	0.161
• Cluster 4	64/188	0.07	0.033	0.01, 0.14		0.027	0.072
Age	188	0.01	0.003	0.00, 0.01		**0.008**	0.032
Season of inclusion **					0.097		

• Spring	47/188	-0.06	0.033	-0.13, 0.00		0.069	0.139
• Summer	37/188	-0.03	0.035	-0.10, 0.04		0.345	0.394
• Winter	47/188	0.02	0.032	-0.04, 0.09		0.464	0.464
Library size	188	1×10^{-6}	7×10^{-6}	0.00, 0.00		0.145	0.193
$R^2 = 0.135$; adjusted $R^2 = 0.096$, AIC = -146.0, overall *P*-value = 9.3×10^{-4}							

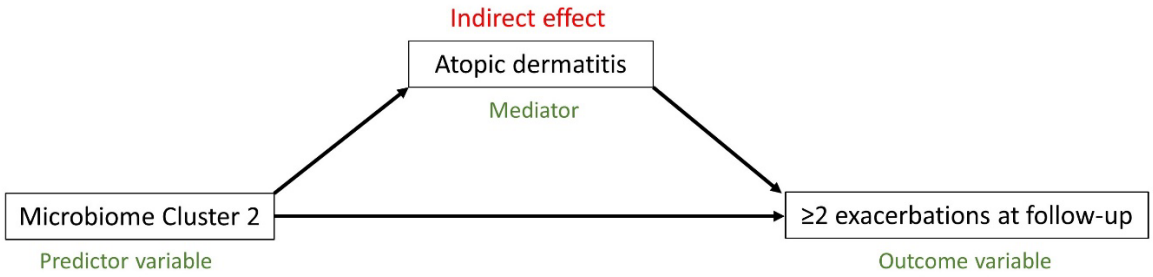
β: beta regression co-efficient, AIC: Akaike information criterion CI: confidence interval, NCS: nasal corticosteroids, SE: standard error, and TGF-β: transforming growth factor β. † Only children with available blood transcriptome data (n=188) were included in the regression analysis. # Variables in the model were selected according to a stepwise process. ¶ Cluster 1 (n=40/188, with the lowest enrichment scores median value) was the reference group for comparison in the stepwise process. * Inclusion center C (n=37/188) was randomly selected as a reference group for comparison in the stepwise process. ** Fall season (n=57/188) was randomly selected as a reference group for comparison in the stepwise process. ‡ The overall *P*-values were computed for taxa-driven clusters, in which all clusters were taken into account (without a reference group selection). The microbiota clusters with the statistically significant q-values after adjusting for covariates are highlighted in bold-face. *q*-value < 0.05 is considered statistically significant.

323 **Table E6:** A stepwise multiple logistic regression model of having ≥ 2 exacerbations/year during follow-
324 up.

≥ 2 exacerbations/year during follow-up							
Characteristic#	**n†**	**OR**	**SE**	**95% CI**	**Overall *P*-value‡**	***P*-value**	***q*-value**
Taxa-driven clusters¶					**1.92 x 10^{-4}**		
• Cluster 1	24/124	0.63	0.667	0.17, 2.33		0.481	0.569
• **Cluster 3**	**11/124**	**0.06**	**1.06**	**0.01, 0.43**		**0.009**	**0.030**
• **Cluster 4**	**41/124**	**0.11**	**0.623**	**0.03, 0.34**		**3.6 x 10^{-4}**	**0.005**
Age	124	1.10	0.054	0.99, 1.23		0.077	0.144
Atopic dermatitis diagnosis	84/124	6.42	0.612	2.05, 23.1		0.002	0.016
Centre of Inclusion*					0.003		
• Centre A	10/124	0.23	1.24	0.01, 2.06		0.242	0.314
• Centre B	11/124	38.4	1.44	3.05, 1,169		0.012	0.030
• Centre D	3/124	0.48	1.48	0.02, 8.43		0.616	0.668
• Centre E	25/124	4.17	0.781	0.96, 21.0		0.068	0.144
• Centre F	27/124	0.93	0.775	0.20, 4.33		0.925	0.925
• Centre G	20/124	3.78	0.811	0.80, 19.8		0.101	0.150
Antibiotic prescription	39/124	5.52	0.622	1.72, 20.2		0.006	0.026
Sequencing run (run 2)	40/124	2.60	0.587	0.85, 8.69		0.104	0.150
	McFadden's R^2 = 0.341, Nagelkerke R^2 = 0.503, AIC = 141.2						

325 AIC: Akaike information criterion, CI: confidence interval, SE: standard error, and OR: odds ratio. † Only children
326 with severe asthma or severe wheezing were followed-up and have future exacerbation data (n=125/241), with
327 n=1/125 patient with missing atopic dermatitis diagnosis was excluded from the regression analysis. # Variables in
328 the model were selected according to a stepwise process. ¶ Cluster 2 (n = 48/124, with the highest % of ≥ 2
329 exacerbations/year during follow-up) was the reference group for comparison in the stepwise process. * Inclusion
330 center C (n=28/124) was randomly selected as a reference group for comparison in the stepwise process. ‡ The
331 overall *P*-values were computed for variables with >2 categories, in which all the variable's categories were taken
332 into account (without a reference group selection). The microbiota clusters with the statistically significant q-values
333 after adjusting for covariates are highlighted in bold-face. *q*-value < 0.05 is considered statistically significant.

334 **Table E7:** Mediation analysis investigating whether atopic dermatitis is mediating the association
335 between microbiota Cluster 2 and follow-up exacerbations.



Mediation analysis	**Unadjusted**			**Adjusted**		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
Average Causal Mediated Effect (ACME)	0.027	-0.011, 0.090	0.154	0.019	-0.021, 0.070	0.382
Average Direct Effect (ADE)	0.258	0.097, 0.410	0.002	0.265	0.115, 0.400	0.002
Total effect	0.285	0.115, 0.440	0.002	0.284	0.124, 0.430	<0.001
Average proportion mediated	0.08	-0.046, 0.340	0.154	0.059	-0.112, 0.270	0.382

336 95% CI: 95% confidence interval. 95% CIs were computed by Quasi-Bayesian approximation (1000 simulations).
337

338

339 **Supplementary Figures Legends:**

340 **Figure E1:** The flow diagram of the inclusion numbers of the U-BIOPRED pediatric cohort, for which the
341 oropharyngeal microbiota were characterized.

342 **Figure E2:** The general data analysis overview performed in this study.

343 **Figure E3:** Directed acyclic graph (DAG) to define the covariates (potential confounders) that may
344 influence the microbiome clusters and the clinical asthma characteristics of interest.

345 **Figure E4:** Analysis of the negative and positive (mock) control samples with respect to the participants'
346 oropharyngeal swabs. A; Distribution of the sample microbiome read counts across the control samples
347 and oropharyngeal swabs. B; A box-whisker plot showing that negative extraction control samples had
348 statistically significant lower Shannon α-diversity measure compared with the participants'
349 oropharyngeal swab samples, in which the P-value was generated by a Mann-Whitney U test. C; Mean
350 relative abundances of most abundant bacterial genera stratified by sample type, in which negative
351 extraction controls mainly showed presence of *Corynebacterium_1*, *Pelomonas*, *Acidibacter* bacterial
352 genera that were not found in oropharyngeal swab samples. D; The positive (mock community) control
353 sample accurately represented all 17 different bacterial genera that were present in a mixture as
354 described in the reference standard (Genomic DNA from Microbial Mock Community B (Even, Low
355 Concentration), v5.1L, for 16S rRNA Gene Sequencing, HM-782D; BEI Resources, NIAID, NIH as part of the
356 Human Microbiome Project).

357 **Figure E5:** Mean relative abundances of most abundant oropharyngeal bacterial genera across all
358 samples (n=241).

359 **Figure E6:** Optimum number of hierarchical clusters based on multiple statistical indices calculated using
360 the Bray-Curtis dissimilarity microbiome measure. Four was the suggested number of clusters based on
361 the majority vote (3 out of 4 tested indices).

362 **Figure E7: A;** Topological data analysis (TDA) network based on the multi-dimensional scaling (MDS)
363 lenses 1 and 2, wherein nodes were colored in accordance with spatial analysis of functional enrichment
364 (SAFE) scores of hierarchical clustering assignments of participants. Four distinct regions (clusters) were
365 observed in the TDA graph showing concordance of the TDA network with the hierarchical clusters
366 assignments. **B;** TDA graph based on the principal co-ordinate analysis (PCoA) lenses 1 and 2, wherein
367 nodes were colored in accordance with SAFE scores of hierarchical clustering assignments of
368 participants.

369 **Figure E8:** Non-metric multi-dimensional scaling (NMDS) and principal co-ordinate analysis (PCoA) two-
370 dimensional (first 2 dimensions) plots were depicted based on the Bray-Curtis dissimilarity microbiome
371 measure, and nodes were colored according to the hierarchical clustering assignments. 95% confidence
372 interval ellipses were added to the participants' nodes representing each cluster.

373 **Figure E9:** Violin box-whisker plot of Shannon α-diversity measure between the taxa-driven Clusters. The
374 median diversity was highest for Cluster 4, followed by Cluster 2, and Cluster3, and was lowest in

375 Cluster1. Overall *P*-value was generated by the Kruskal-Wallis H test and pairwise *P*-values were
376 generated by Dunn's post-hoc test.

377 **Figure E10:** Mean relative abundances of most abundant oropharyngeal bacterial stratified by the
378 clusters.

379 **Figure E11: A;** Phylum-level relative abundances for each sample, grouped by the clusters. **B;** Genus-level
380 relative abundances for each sample, grouped by the clusters. Only the most abundant bacterial taxa
381 (≥1% and ≥5% relative abundances for the phylum and genus levels, respectively) are shown.

382 **Figure E12:** Topological data analysis (TDA) network based on the multi-dimensional scaling (MDS) lenses
383 1 and 2, wherein nodes were colored according to the relative abundance of most abundant and
384 differential amplicon sequence variants (ASVs); *Streptococcus* (**A**), *Veillonella* (**B**), *Rothia* (**C**), and
385 *Haemophilus* (**D**). TDA regions showing the highest relative abundances of respective bacterial taxa are
386 concordant with the clustering assignments regions shown in Figure E7A.

387 **Figure E13:** Topological data analysis (TDA) network based on the principal coordinate analysis (PCoA)
388 lenses 1 and 2, wherein nodes were colored according to the relative abundance of most abundant and
389 differential amplicon sequence variants (ASVs); *Streptococcus* (**A**), *Veillonella* (**B**), *Rothia* (**C**), and
390 *Haemophilus* (**D**). TDA regions showing the highest relative abundances of respective bacterial taxa are
391 concordant with the clustering assignments regions shown in Figure E7B.

392 **Figure E14:** Decision tree computed from a training data set (≈75% random participants) showed that the
393 relative abundance of only 3 main bacterial genera could stratify participants into corresponding
394 clusters. The performance of the decision tree was tested on a left-out validation set (≈25% random
395 participants) exhibiting an accuracy of correct cluster classification = 0.86 (95% CI: 0.75-0.94, *P*-value =
396 8.24×10^{-15}).

397 **Figure E15:** Violin box-and-whisker plots showing the forced expiratory volume in 1 second (FEV_1) %
398 predicted values pre- and post-salbutamol between the taxa-driven clusters. Overall *P*-values were
399 generated by the Kruskal-Wallis H test and pairwise *P*-values were generated by Dunn's post-hoc test.

400 **Figure E16:** Violin box-and-whisker plots showing the number of aeroallergens sensitization between the
401 taxa-driven clusters. Overall *P*-value was generated by the Kruskal-Wallis H test and pairwise *P*-values
402 were generated by Dunn's post-hoc test.

403 **Figure E17:** Summary overview of clinical features associated with the taxa-driven clusters.

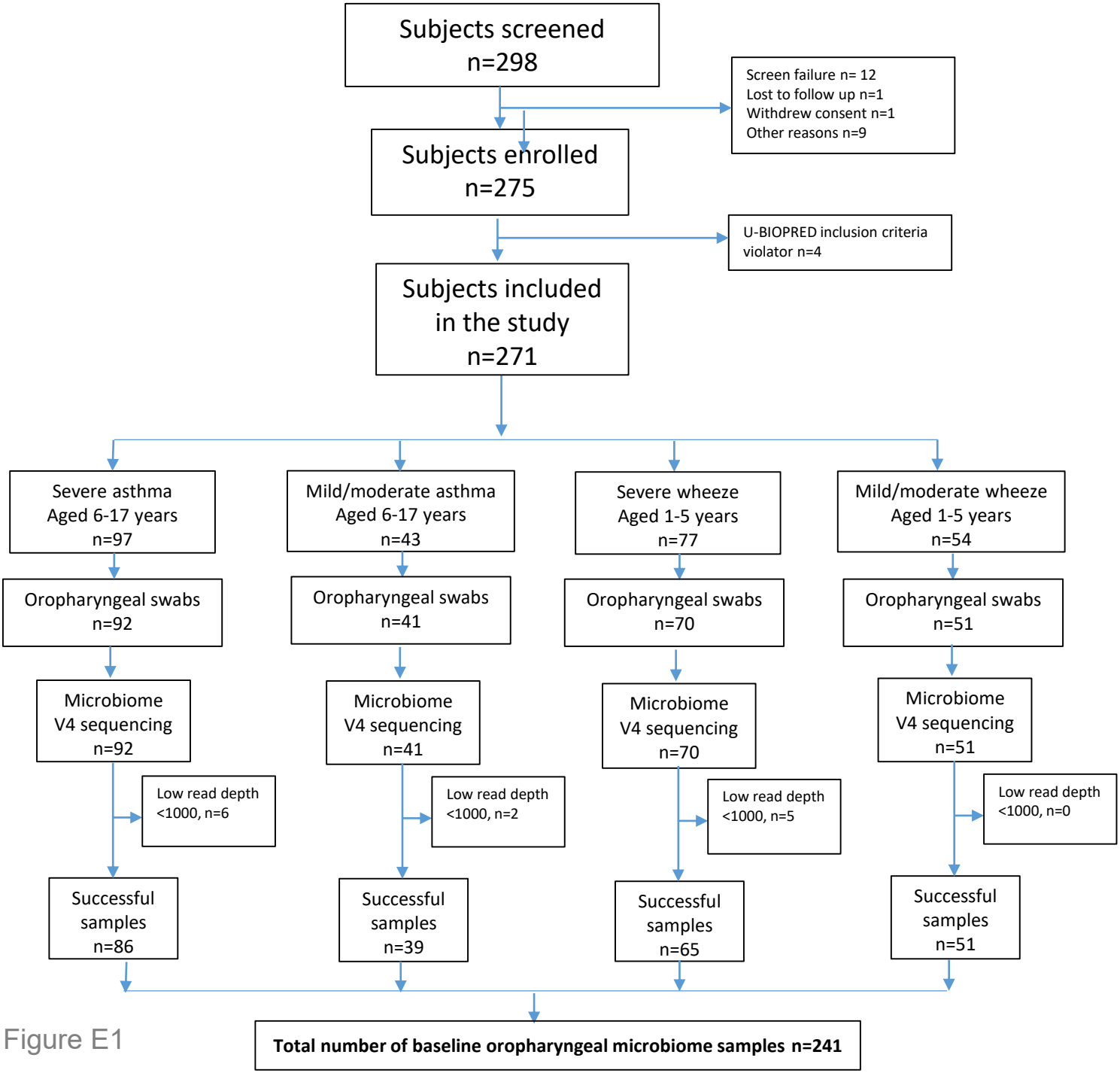

Subjects screened
n=298
Screen failure n= 12
Lost to follow up n=1
Withdrew consent n=1
Other reasons n=9
Subjects enrolled
n=275
U-BIOPRED inclusion criteria violator n=4
Subjects included in the study
n=271
Severe asthma
Aged 6-17 years
n=97
Mild/moderate asthma
Aged 6-17 years
n=43
Severe wheeze
Aged 1-5 years
n=77
Mild/moderate wheeze
Aged 1-5 years
n=54
Oropharyngeal swabs
n=92
Oropharyngeal swabs
n=41
Oropharyngeal swabs
n=70
Oropharyngeal swabs
n=51
Microbiome
V4 sequencing
n=92
Microbiome
V4 sequencing
n=41
Microbiome
V4 sequencing
n=70
Microbiome
V4 sequencing
n=51
Low read depth
<1000, n=6
Low read depth
<1000, n=2
Low read depth
<1000, n=5
Low read depth
<1000, n=0
Successful samples
n=86
Successful samples
n=39
Successful samples
n=65
Successful samples
n=51
Total number of baseline oropharyngeal microbiome samples n=241

Figure E1

Figure E2

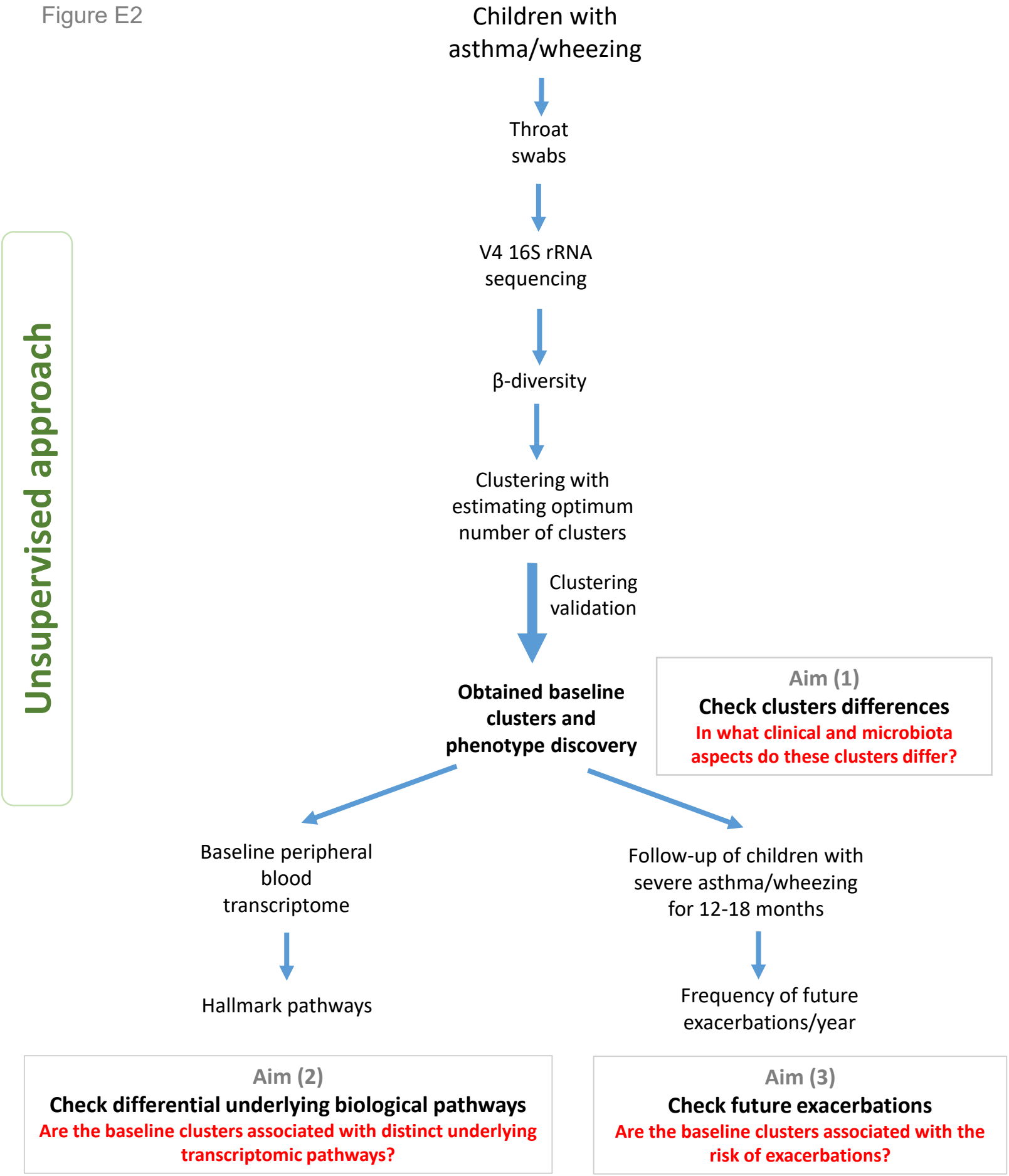
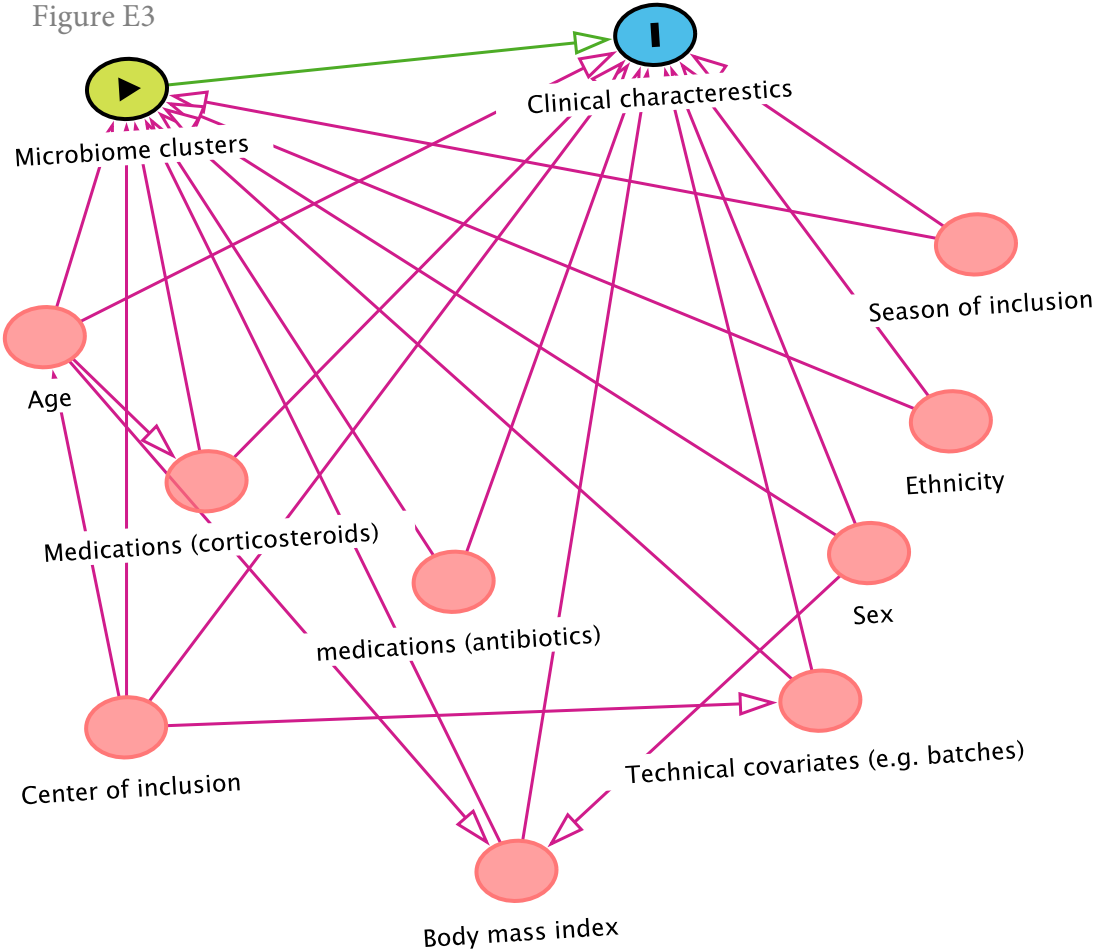

Children with asthma/wheezing
Throat swabs
V4 16S rRNA sequencing
β-diversity
Clustering with estimating optimum number of clusters
Clustering validation
Obtained baseline clusters and phenotype discovery
Unsupervised approach
Aim (1)
Check clusters differences
In what clinical and microbiota aspects do these clusters differ?
Baseline peripheral blood transcriptome
Hallmark pathways
Follow-up of children with severe asthma/wheezing for 12-18 months
Frequency of future exacerbations/year
Aim (2)
Check differential underlying biological pathways
Are the baseline clusters associated with distinct underlying transcriptomic pathways?
Aim (3)
Check future exacerbations
Are the baseline clusters associated with the risk of exacerbations?

Figure E3


Clinical characterestics
Microbiome clusters
Season of inclusion
Age
Ethnicity
Medications (corticosteroids)
Sex
medications (antibiotics)
Center of inclusion
Technical covariates (e.g. batches)
Body mass index

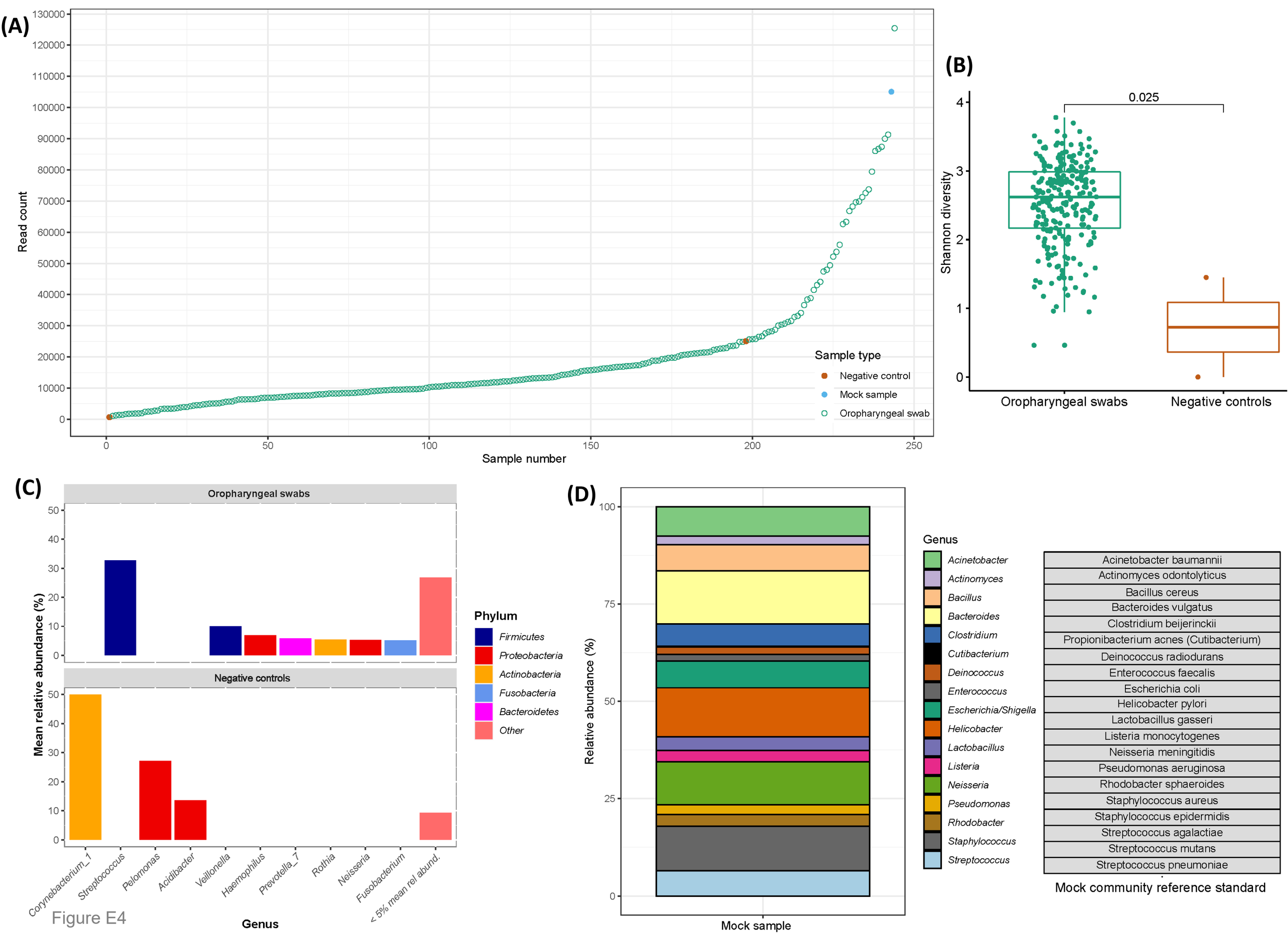

(A)
Read count
Sample number
Sample type
Negative control
Mock sample
Oropharyngeal swab
(B)
Shannon diversity
0.025
Oropharyngeal swabs
Negative controls
(C)
Oropharyngeal swabs
Negative controls
Mean relative abundance (%)
Phylum
Firmicutes
Proteobacteria
Actinobacteria
Fusobacteria
Bacteroidetes
Other
Corynebacterium_1
Streptococcus
Pelomonas
Acidibacter
Veillonella
Haemophilus
Prevotella_7
Rothia
Neisseria
Fusobacterium
< 5% mean rel abund.
Genus
(D)
Relative abundance (%)
Mock sample
Genus
Acinetobacter
Actinomyces
Bacillus
Bacteroides
Clostridium
Cutibacterium
Deinococcus
Enterococcus
Escherichia/Shigella
Helicobacter
Lactobacillus
Listeria
Neisseria
Pseudomonas
Rhodobacter
Staphylococcus
Streptococcus
Acinetobacter baumannii
Actinomyces odontolyticus
Bacillus cereus
Bacteroides vulgatus
Clostridium beijerinckii
Propionibacterium acnes (Cutibacterium)
Deinococcus radiodurans
Enterococcus faecalis
Escherichia coli
Helicobacter pylori
Lactobacillus gasseri
Listeria monocytogenes
Neisseria meningitidis
Pseudomonas aeruginosa
Rhodobacter sphaeroides
Staphylococcus aureus
Staphylococcus epidermidis
Streptococcus agalactiae
Streptococcus mutans
Streptococcus pneumoniae
Mock community reference standard

Figure E4

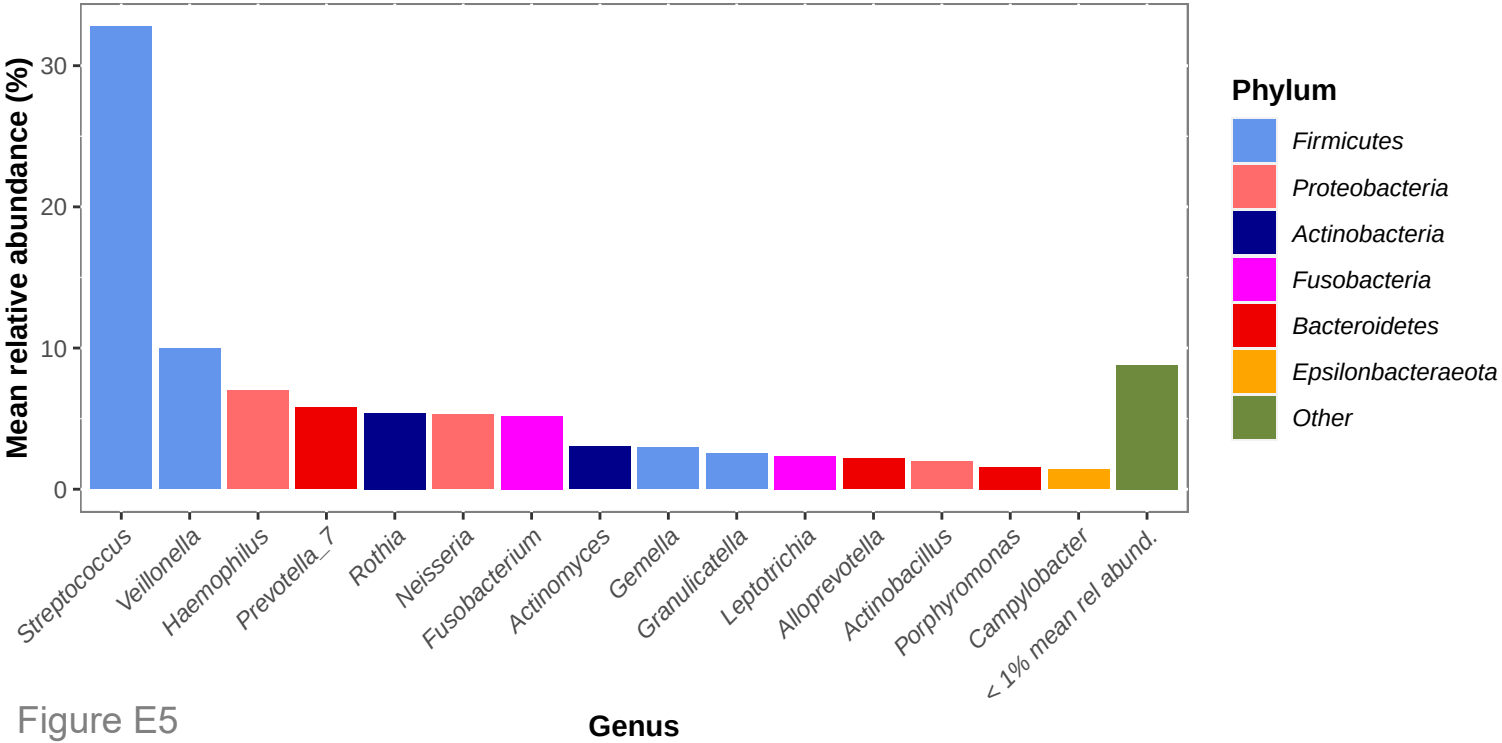

Mean relative abundance (%)
30
20
10
0
Streptococcus
Veillonella
Haemophilus
Prevotella_7
Rothia
Neisseria
Fusobacterium
Actinomyces
Gemella
Granulicatella
Leptotrichia
Alloprevotella
Actinobacillus
Porphyromonas
Campylobacter
< 1% mean rel abund.
Genus
Phylum
Firmicutes
Proteobacteria
Actinobacteria
Fusobacteria
Bacteroidetes
Epsilonbacteraeota
Other

Figure E5

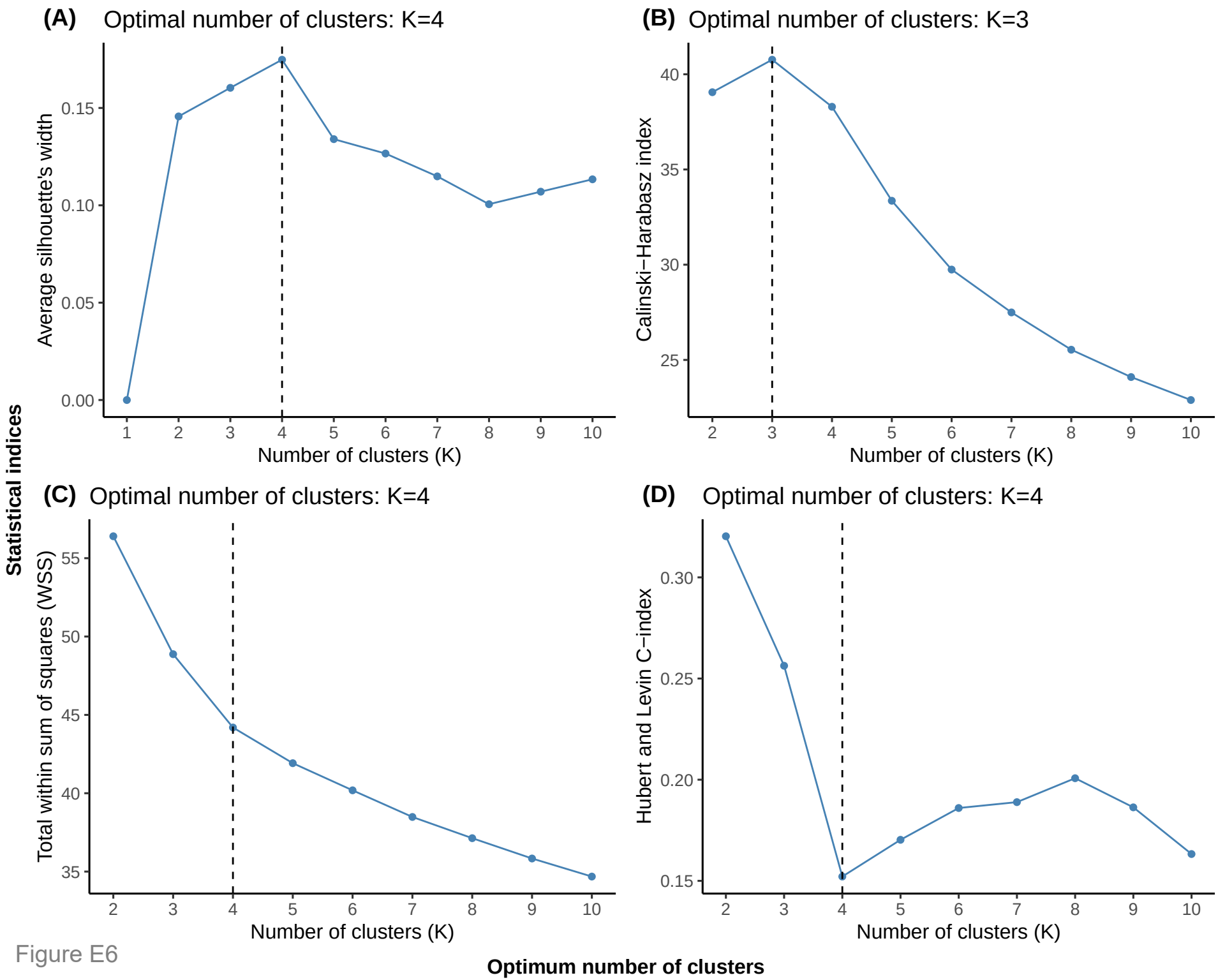

(A) Optimal number of clusters: K=4
Average silhouette's width
0.00
0.05
0.10
0.15
1 2 3 4 5 6 7 8 9 10
Number of clusters (K)
(B) Optimal number of clusters: K=3
Calinski−Harabasz index
25
30
35
40
2 3 4 5 6 7 8 9 10
Number of clusters (K)
(C) Optimal number of clusters: K=4
Total within sum of squares (WSS)
35
40
45
50
55
2 3 4 5 6 7 8 9 10
Number of clusters (K)
(D) Optimal number of clusters: K=4
Hubert and Levin C−index
0.15
0.20
0.25
0.30
2 3 4 5 6 7 8 9 10
Number of clusters (K)
Statistical indices
Optimum number of clusters

Figure E6

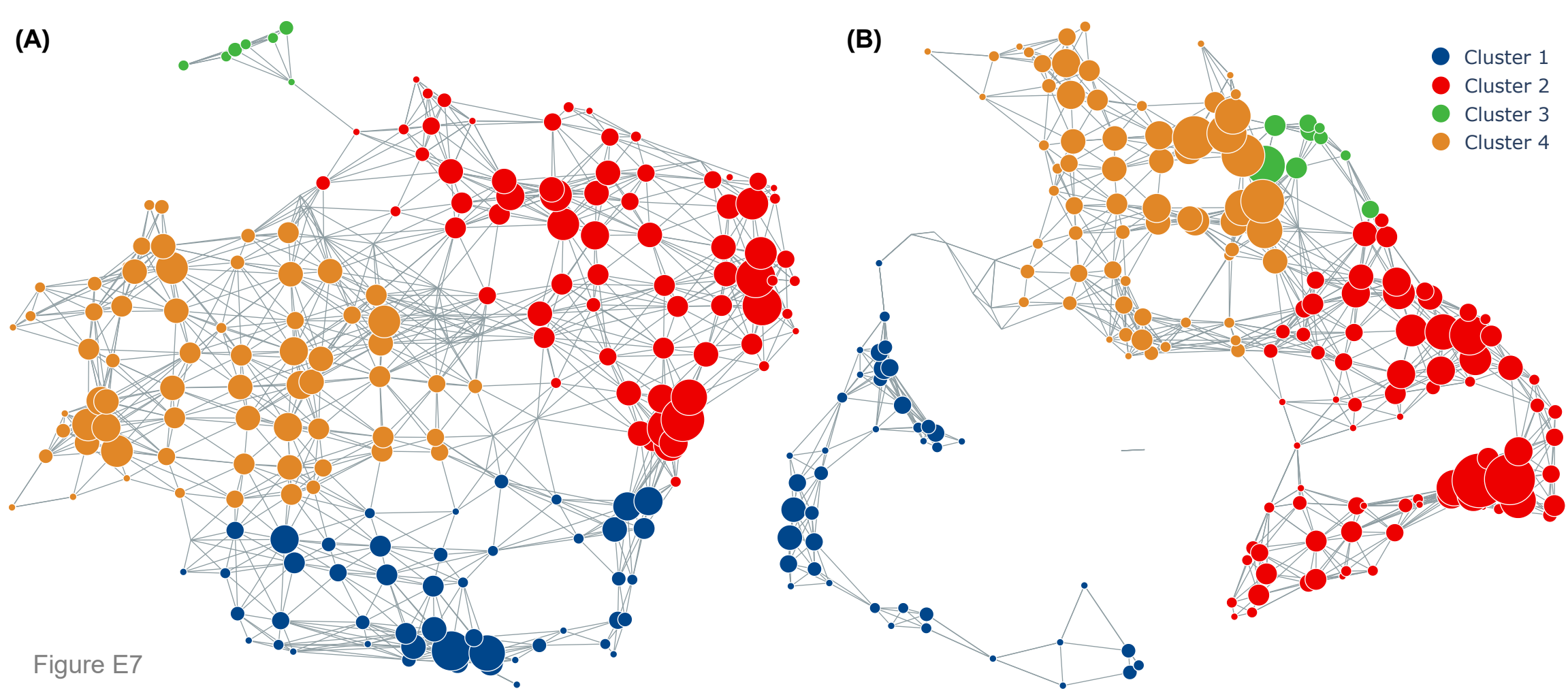
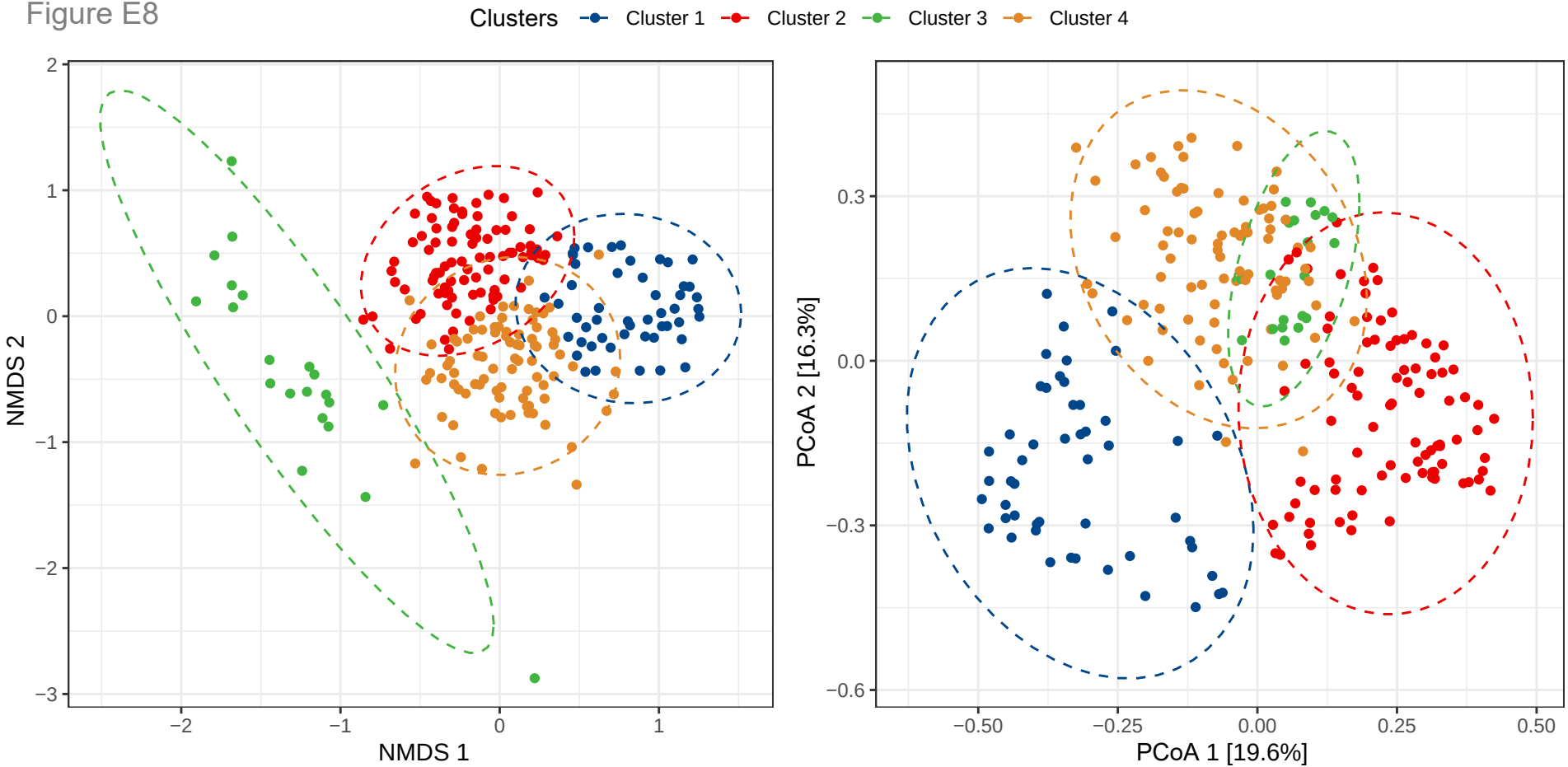


Figure E7

Figure E8


Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
NMDS 2
NMDS 1
2
1
0
−1
−2
−3
−2
−1
0
1
PCoA 2 [16.3%]
PCoA 1 [19.6%]
0.3
0.0
−0.3
−0.6
−0.50
−0.25
0.00
0.25
0.50

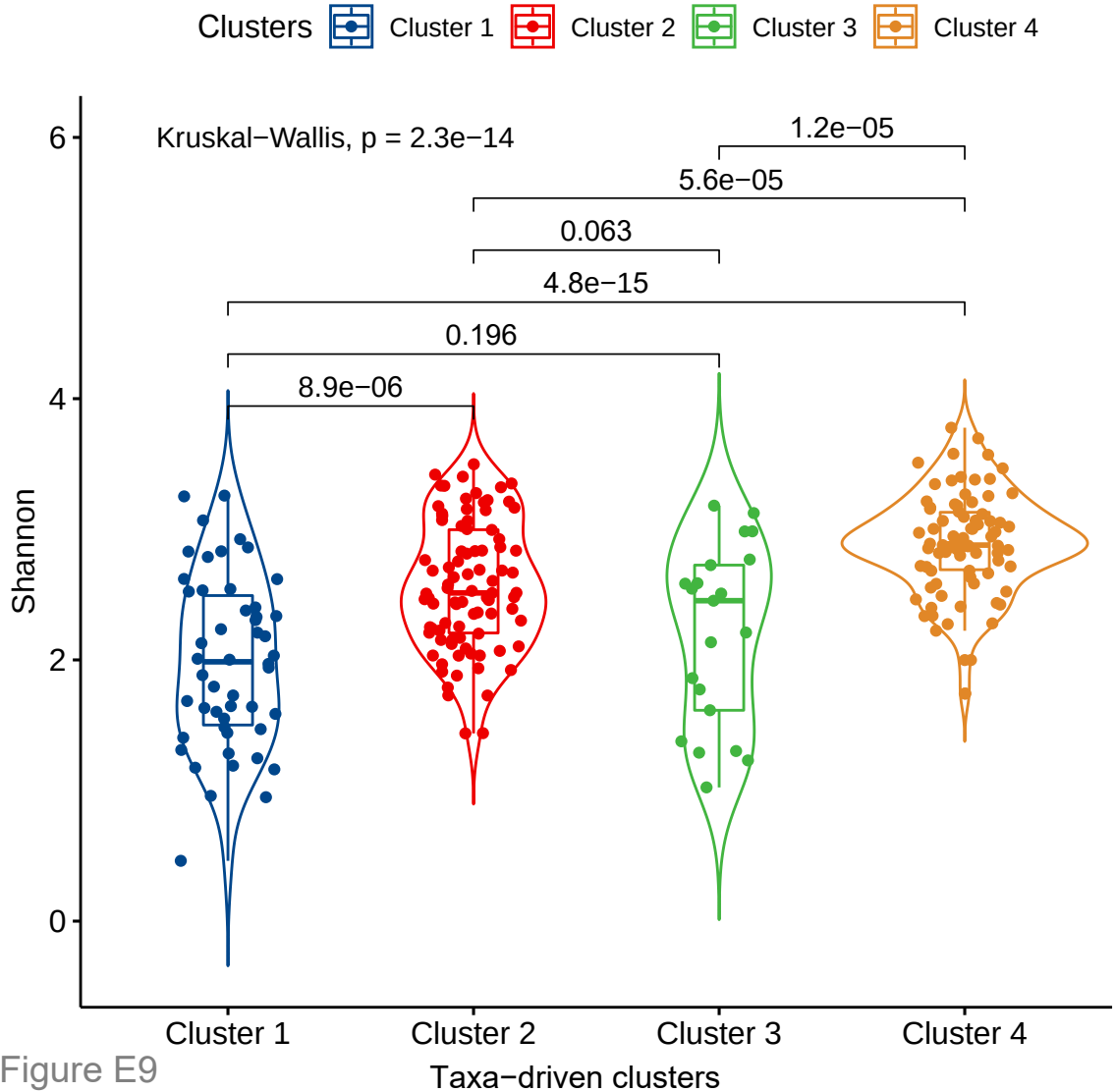

Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Kruskal−Wallis, p = 2.3e−14
1.2e−05
5.6e−05
0.063
4.8e−15
0.196
8.9e−06
Shannon
6
4
2
0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Taxa−driven clusters

Figure E9

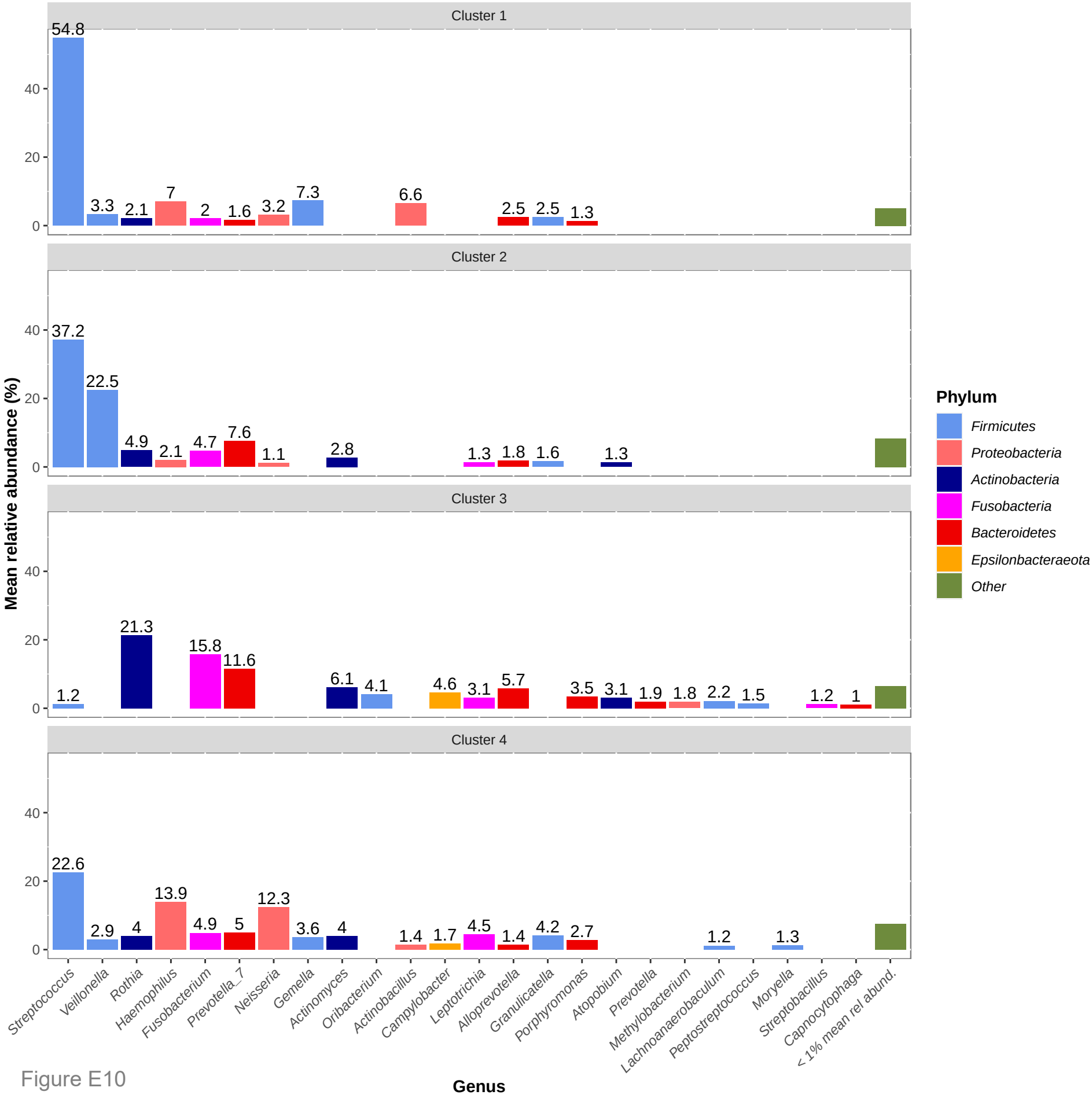

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Mean relative abundance (%)
0
20
40
54.8
3.3
2.1
7
2
1.6
3.2
7.3
6.6
2.5
2.5
1.3
37.2
22.5
4.9
2.1
4.7
7.6
1.1
2.8
1.3
1.8
1.6
1.3
1.2
21.3
15.8
11.6
6.1
4.1
4.6
3.1
5.7
3.5
3.1
1.9
1.8
2.2
1.5
1.2
1
22.6
2.9
4
13.9
4.9
5
12.3
3.6
4
1.4
1.7
4.5
1.4
4.2
2.7
1.2
1.3
Streptococcus
Veillonella
Rothia
Haemophilus
Fusobacterium
Prevotella_7
Neisseria
Gemella
Actinomyces
Oribacterium
Actinobacillus
Campylobacter
Leptotrichia
Alloprevotella
Granulicatella
Porphyromonas
Atopobium
Prevotella
Methylobacterium
Lachnoanaerobaculum
Peptostreptococcus
Moryella
Streptobacillus
Capnocytophaga
< 1% mean rel abund.
Genus
Phylum
Firmicutes
Proteobacteria
Actinobacteria
Fusobacteria
Bacteroidetes
Epsilonbacteraeota
Other

Figure E10


(A)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Relative abundance (%)
100
75
50
25
0
Sample
Phylum
Firmicutes
Actinobacteria
Proteobacteria
Bacteroidetes
Fusobacteria
Patescibacteria
Cyanobacteria
Epsilonbacteraeota
Spirochaetes
Synergistetes
Acidobacteria
< 1% relative abundance
(B)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Relative abundance (%)
100
75
50
25
0
Sample
Genus
Streptococcus
Rothia
Veillonella
Actinobacillus
Haemophilus
Fusobacterium
Gemella
Neisseria
Prevotella_7
Methylobacterium
Actinomyces
Leptotrichia
Oribacterium
unclassified
Porphyromonas
Alloprevotella
Moryella
Atopobium
Ralstonia
Bifidobacterium
Streptobacillus
Peptostreptococcus
Pseudomonas
Campylobacter
Granulicatella
Lautropia
Selenomonas_3
Mucilaginibacter
Prevotella
Lactobacillus
Lachnoanaerobaculum
Capnocytophaga
< 5% relative abundance

Figure E11

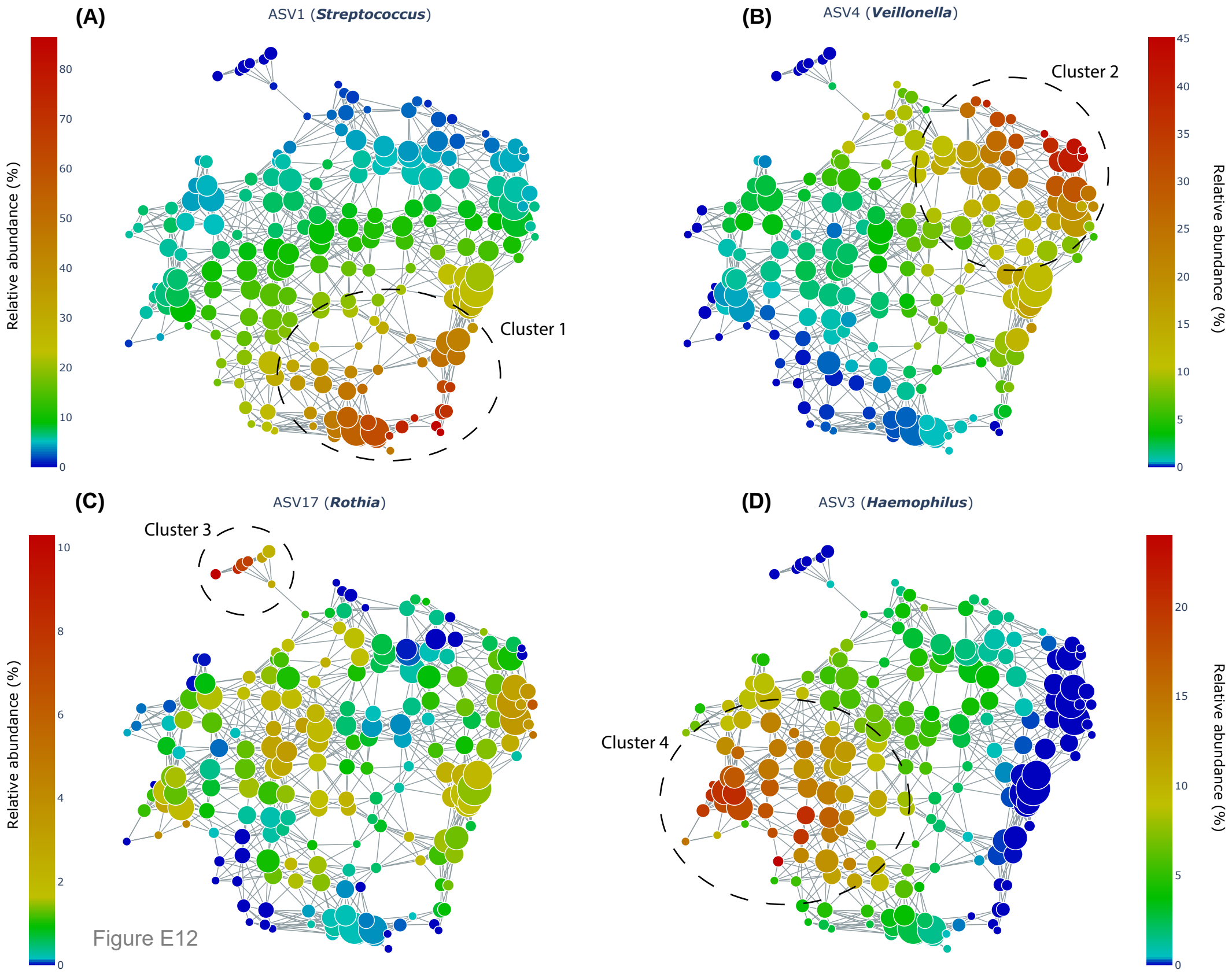

(A)
ASV1 (Streptococcus)
Relative abundance (%)
0
10
20
30
40
50
60
70
80
Cluster 1
(B)
ASV4 (Veillonella)
Cluster 2
Relative abundance (%)
0
5
10
15
20
25
30
35
40
45
(C)
ASV17 (Rothia)
Cluster 3
Relative abundance (%)
0
2
4
6
8
10
(D)
ASV3 (Haemophilus)
Cluster 4
Relative abundance (%)
0
5
10
15
20

Figure E12

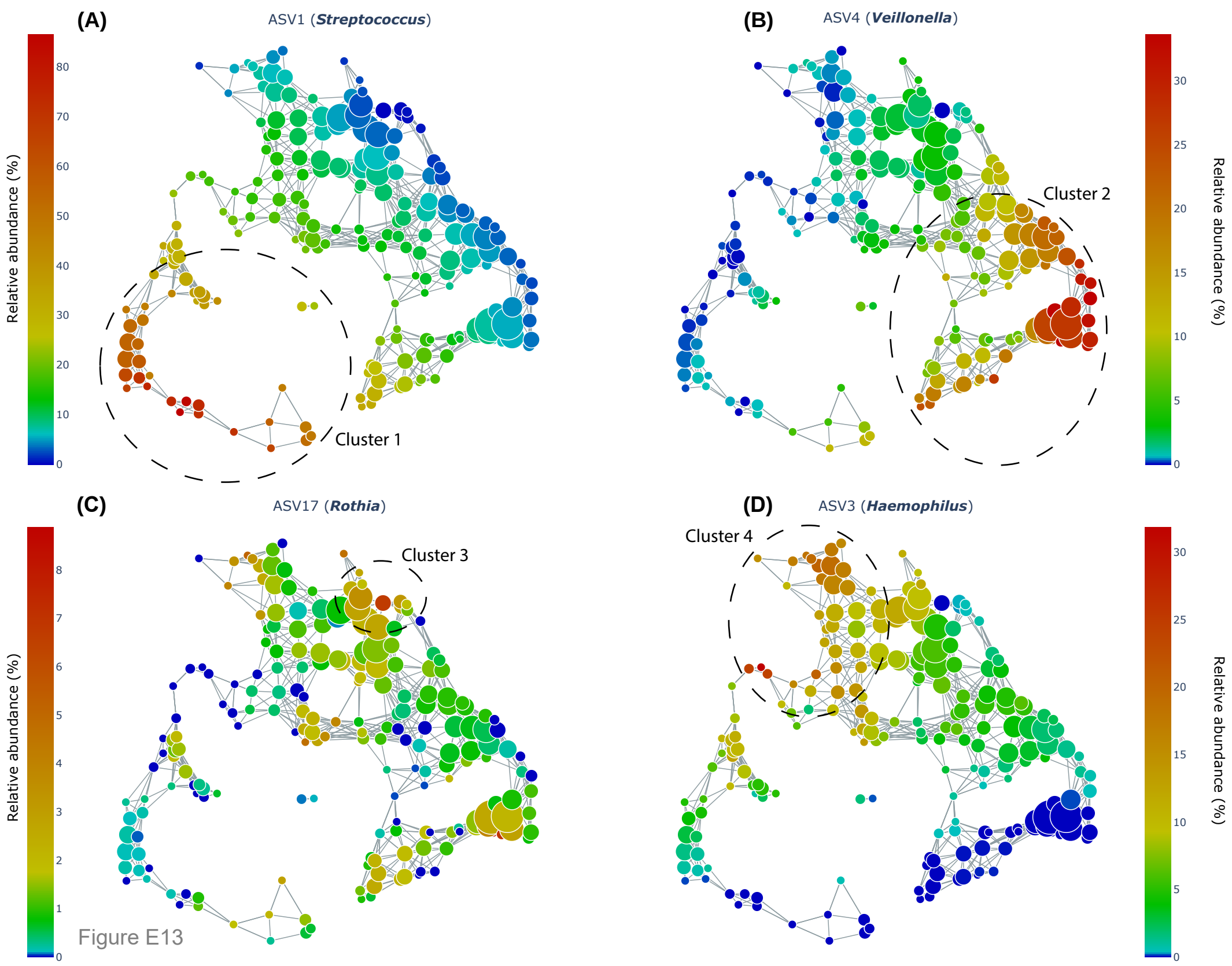
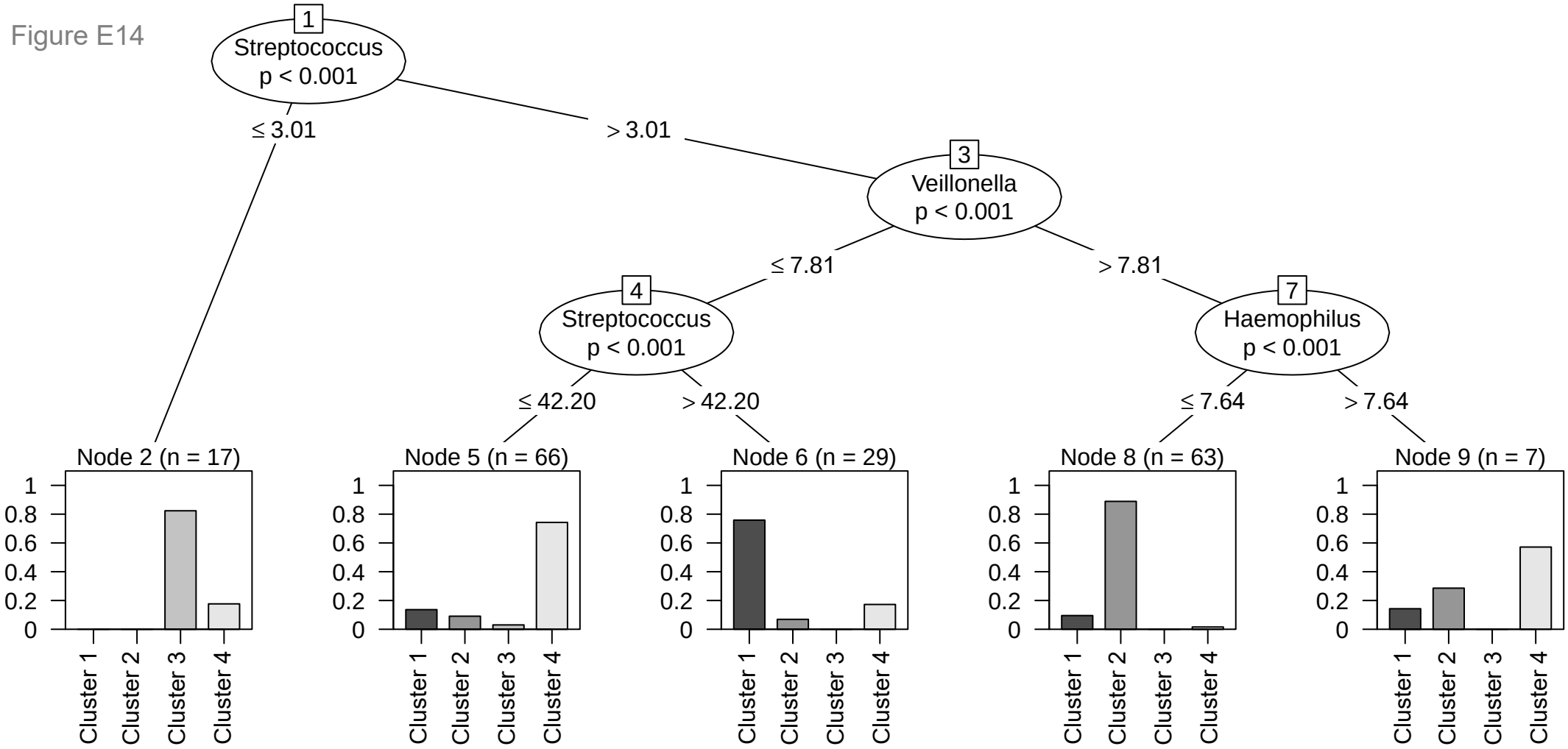

(A)
ASV1 (Streptococcus)
Relative abundance (%)
0
10
20
30
40
50
60
70
80
Cluster 1
(B)
ASV4 (Veillonella)
Cluster 2
Relative abundance (%)
0
5
10
15
20
25
30
(C)
ASV17 (Rothia)
Cluster 3
Relative abundance (%)
0
1
2
3
4
5
6
7
8
(D)
ASV3 (Haemophilus)
Cluster 4
Relative abundance (%)
0
5
10
15
20
25
30

Figure E13

Figure E14


1
Streptococcus
p < 0.001
≤ 3.01
> 3.01
3
Veillonella
p < 0.001
≤ 7.81
> 7.81
4
Streptococcus
p < 0.001
7
Haemophilus
p < 0.001
≤ 42.20
> 42.20
≤ 7.64
> 7.64
Node 2 (n = 17)
Node 5 (n = 66)
Node 6 (n = 29)
Node 8 (n = 63)
Node 9 (n = 7)
1
0.8
0.6
0.4
0.2
0
Cluster 1
Cluster 2
Cluster 3
Cluster 4

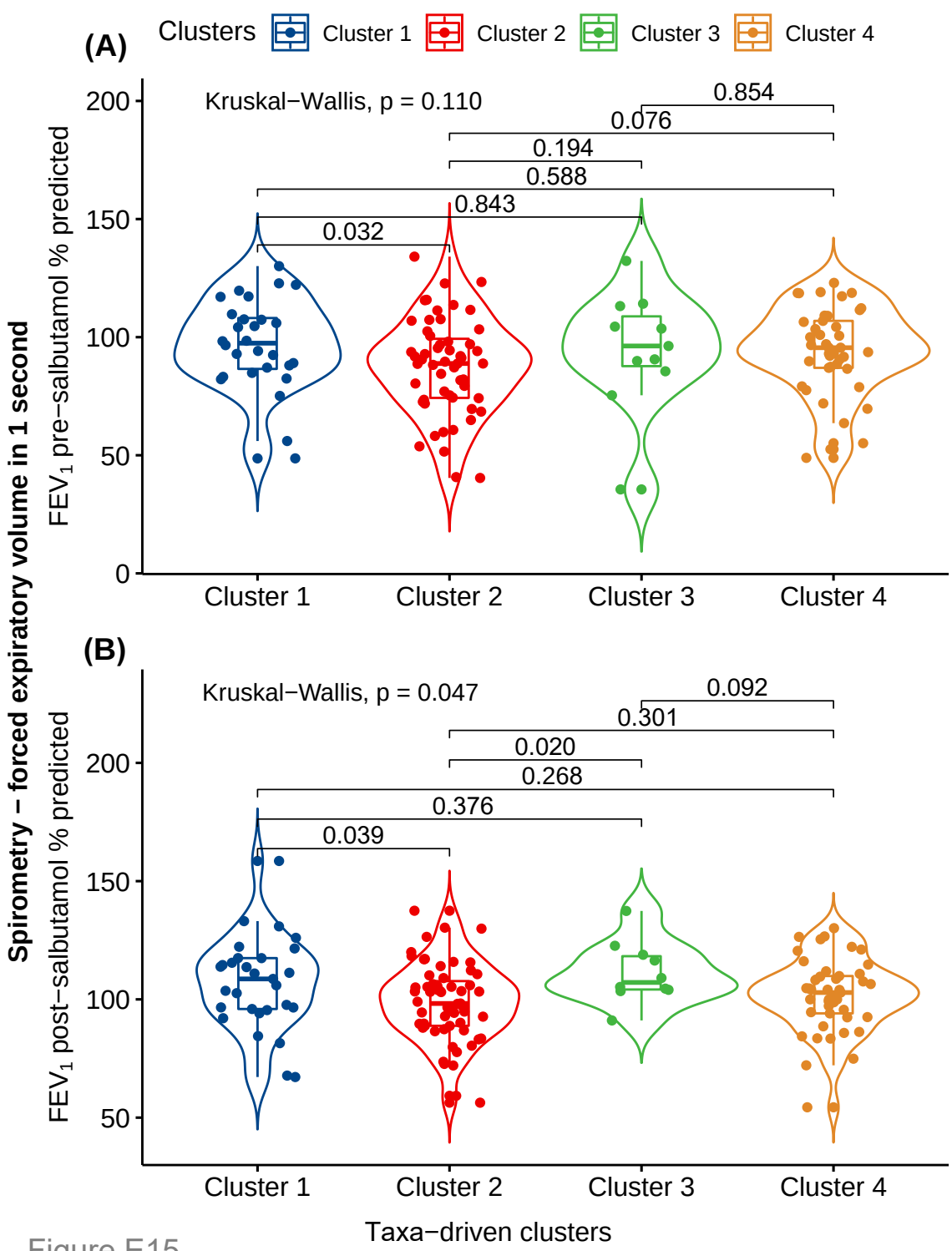
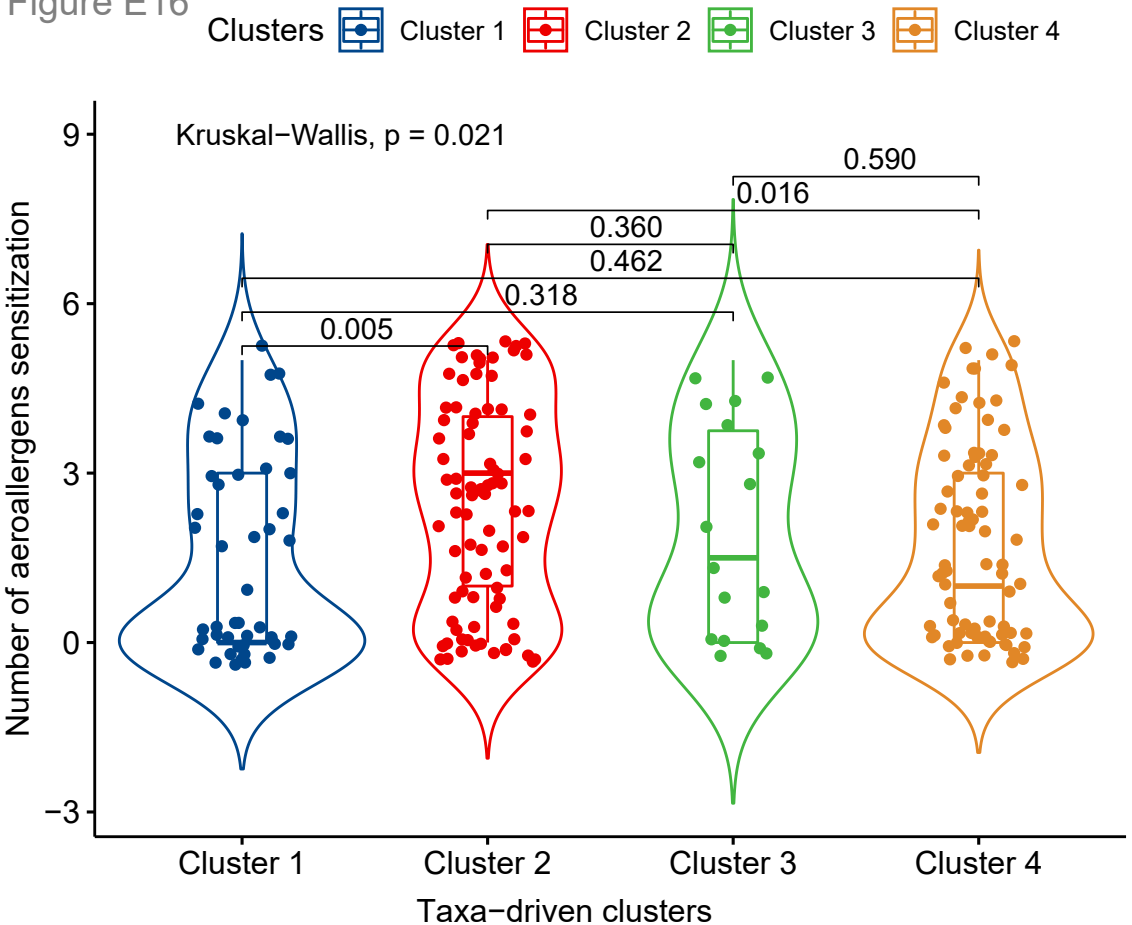
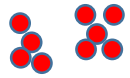

Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
(A)
Kruskal−Wallis, p = 0.110
0.854
0.076
0.194
0.588
0.843
0.032
FEV1 pre−salbutamol % predicted
200
150
100
50
0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
(B)
Kruskal−Wallis, p = 0.047
0.092
0.301
0.020
0.268
0.376
0.039
FEV1 post−salbutamol % predicted
200
150
100
50
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Taxa−driven clusters
Spirometry − forced expiratory volume in 1 second

Figure E15

Figure E16


Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Kruskal−Wallis, p = 0.021
0.590
0.016
0.360
0.462
0.318
0.005
9
6
3
0
−3
Number of aeroallergens sensitization
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Taxa−driven clusters



Veillonella-dominant
↑ % children with severe school-age asthma
↑ Atopic dermatitis, and ↑ Atopy
↓ FEV_1 % predicted post-bronchodilator
↑ TGF-β transcriptomic pathway
↑ % of children with severe symptoms who had ≥2 future exacerbations

Streptococcus-dominant
↓ Atopic dermatitis, and ↓ Atopy
↓ Wnt/β-Catenin transcriptomic pathway
↓ TGF-β transcriptomic pathway

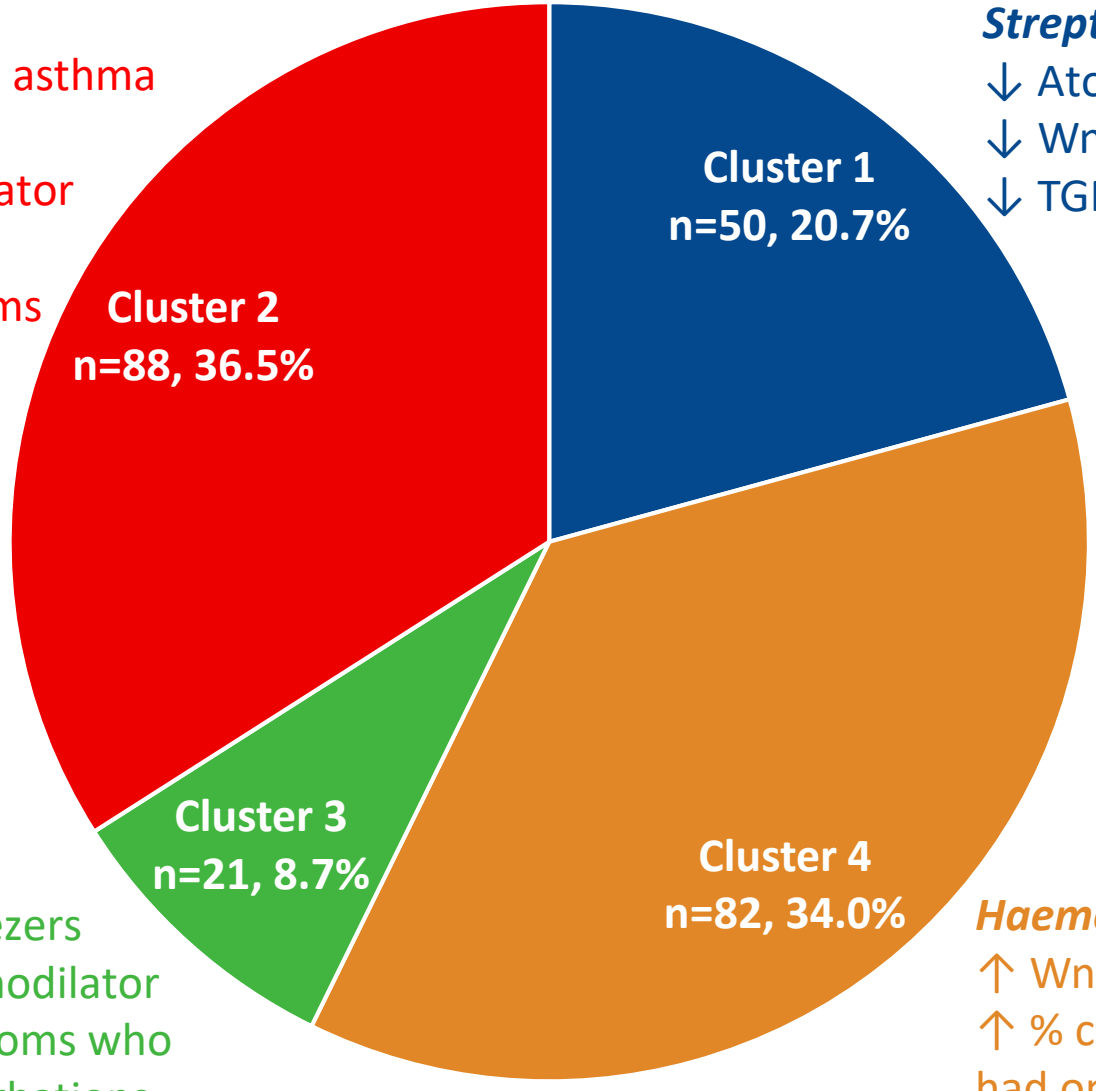
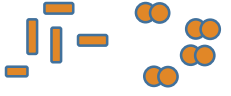


Figure E17

Rothia-dominant
↑ % severe preschool-age wheezers
↑ FEV_1 % predicted post-bronchodilator
↑ % children with severe symptoms who did not experience future exacerbations



Haemophilus/Neisseria-dominant
↑ Wnt/β-Catenin transcriptomic pathway
↑ % children with severe symptoms who had one future exacerbation

U-BIOPRED

The U-BIOPRED consortium wishes to acknowledge the help and expertise of the following individuals and groups without whom, the study would not have been possible:

Definition for the U-BIOPRED Study Group Supplementary authors
Clinical site research leads
Platform leads
Data cleaning team
Data analysis team
Scientific Board and Management Board members
Core project management staff

Definition for the U-BIOPRED Study Group Contributors list
Significant involvement in the clinical study

Use of list:
This list is to be used for all non-core clinical papers.

Instructions
Follow up clinical papers should re-use the baseline cohort description paper lists, in order to recognize the clinical staff involved in the study.

U-BIOPRED Supplementary authors	
Name	**Affiliation**
Abdel-Aziz MI	Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
Adcock I M	National Heart and Lung Institute, Imperial College, London, UK;
Andersson LI	Department of Respiratory Medicine, Karolinska University Hospital, Stockholm, Sweden
Auffray C	European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France;
Badi YE,	National Heart and Lung Institute, Imperial College, London, UK;
Bakke P	Department of Clinical Science, University of Bergen, Bergen, Norway;
Bansal A T	Acclarogen Ltd, St. John's Innovation Centre, Cambridge, UK;
Baribaud F	Janssen R&D, LLC, Spring House, PA, USA
Bates S	Respiratory Therapeutic Unit, GSK, London, UK;
Bel E H	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Bigler J	*Previously Amgen Inc*
Billing B	Department of Respiratory Medicine, Karolinska University Hospital, Stockholm, Sweden
Bisgaard H	COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark
Boedigheimer M J	Amgen Inc.; Thousand Oaks, USA
Bønnelykke K	COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark;
Brandsma J	University of Southampton, Southampton, UK
Brinkman P	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Bucchioni E	Chiesi Pharmaceuticals SPA, Parma, Italy
Burg D	Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, Southampton, UK

Bush A	National Heart and Lung Institute, Imperial College, London, UK; Royal Brompton and Harefield NHS trust, UK
Caruso M	Dept. Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy;
Chalekis R	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Chanez P	Assistance publique des Hôpitaux de Marseille - Clinique des bronches, allergies et sommeil, Aix Marseille Université, Marseille, France
Chung F K	National Heart and Lung Institute, Imperial College, London, UK;
Checa T	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Compton C H	Respiratory Therapeutic Unit, GSK, London, UK
Corfield J	Areteva R&D, Nottingham, UK;
Cunoosamy D	Sanofi, Cambridge, USA
D'Amico A	University of Rome 'Tor Vergata', Rome Italy;
Dahlén B	Department of Respiratory Medicine, Karolinska University Hospital & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Dahlén S E	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, and Department of Respiratory Medicine, Karolinska University Hospital, Stockholm, Sweden
De Meulder B	European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France;
Djukanovic R	NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, UK;
Erpenbeck V J	Translational Medicine, Respiratory Profiling, Novartis Institutes for Biomedical Research, Basel, Switzerland;
Erzen D	Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Fichtner K	Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Fleming L J	National Heart and Lung Institute, Imperial College, London, UK; Royal Brompton and Harefield NHS trust, UK
Formaggio E	*Previously CROMSOURCE, Verona Italy*
Fowler S J	Division of infection, immunity and respiratory medicine, School of biological sciences, University of Manchester, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom
Frey U	University Children's Hospital, Basel, Switzerland;
Gahlemann M	Boehringer Ingelheim (Schweiz) GmbH,Basel, Switzerland;
Geiser T	Department of Respiratory Medicine, University Hospital Bern, Switzerland;
Goss V	NIHR Respiratory Biomedical Research Unit, University Hospital Southampton NHS Foundation Trust, Integrative Physiology and Critical Illness Group, Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine, University of Southampton, Southampton, UK;
Guo Y	Data Science Institute, Imperial College, London, UK;
Hashimoto S	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Haughney J	International Primary Care Respiratory Group, Aberdeen, Scotland;
Hedlin G	Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden;
Hekking P W	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Higenbottam T	Allergy Therapeutics, West Sussex, UK;

Hohlfeld J M	Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany
Holweg C	Respiratory and Allergy Diseases, Genentech, San Francisco, USA
Horváth I	Department of Pulmonology and Department of Public Health, Semmelweis University, Budapest, Hungary.
Howarth P	NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences and Human Development and Health, Southampton, UK
James A J	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden;
Knowles RG	Knowles Consulting Ltd, Stevenage. UK;
Kolmert J	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Konradsen J	Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Krug N	Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany;
Lazarinis N	Department of Respiratory Medicine, Karolinska University Hospital & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Li C-X	Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
Loza M J	Janssen R&D, LLC, Spring House, PA, USA
Lutter R	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Manta A	Roche Diagnostics GmbH, Mannheim, Germany
Masefield S	European Lung Foundation, Sheffield, UK;
Maitland-van der Zee Anke-Hilse	Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands;
Matthews J G	Respiratory and Allergy Diseases, Genentech, San Francisco, USA;
Mazein A	European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France
Middelveld R J M	Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Miralpeix M	Almirall, Barcelona, Spain;
Montuschi P	Università Cattolica del Sacro Cuore, Milan, Italy;
Murray C S	Division of infection, immunity and respiratory medicine, School of biological sciences, University of Manchester, Manchester University NHS Foundation Trust, and Manchester Academic Health Science Centre, Manchester, United Kingdom
Musial J	Dept. of Medicine, Jagiellonian University Medical College, Krakow, Poland
Mumby, S	National Heart and Lung Institute, Imperial College, London, UK
Myles D	Respiratory Therapeutic Unit, GSK, London, UK;
Nordlund B	Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Pandis I	Data Science Institute, Imperial College, London, UK
Pavlidis S	National Heart and Lung Institute, Imperial College, London, UK
Postle A	University of Southampton, UK
Powel P	European Lung Foundation, Sheffield, UK;
Praticò G	CROMSOURCE, Verona, Italy
Puig Valls M	CROMSOURCE, Barcelona, Spain
Rao N	Janssen R&D, LLC, Spring House, PA, USA
Reinke S	Institute of Environmental Medicine, Karolinska Institutet, Stockoholm, Sweden
Riley J	Respiratory Therapeutic Unit, GSK, London, UK;
Roberts A	Asthma UK, London, UK;

Roberts G	NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences and Human Development and Health, Southampton, UK;
Rowe A	Janssen R&D, UK;
Sandström T	Dept of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden;
Schofield JPR	Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, Southampton, UK
Seibold W	Boehringer Ingelheim Pharma GmbH, Biberach, Germany
Shaw D E	Respiratory Research Unit, University of Nottingham, UK;
Sigmund R	Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Singer F	Pediatric Respiratory Medicine, Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland.
Skipp P J	Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, Southampton, UK
Smicker M	Sanofi, Cambridge, USA
Sousa A R	Respiratory Therapeutic Unit, GSK, London, UK;
Sparreman-Mikus M	Department of Medicine Huddinge, Karolinska Institutet, and Department of Respiratory Medicine, Karolinska University Hospital, Stockholm, Sweden
Sterk P J	Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Ström M	Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
Sun K	Data Science Institute, Imperial College, London, UK
Thornton B	MSD, USA
Uddin M	AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden
Versi A	National Heart and Lung Institute, Imperial College, London, UK
Vestbo J	Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester and University Hospital of South Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom
Vissing N H	COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark;
Wagers S S	BioSci Consulting, Maasmechelen, Belgium
Wheelock A M	Respiratory Medicine Unit, Department of Medicine Solna and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; and Department of Respiratory Medicine and Allergy, Karolinska University Hospital Solna, Stockholm, Sweden;
Wheelock C E	Institute of Environmental Medicine, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden;
Wilson S J	Histochemistry Research Unit, Faculty of Medicine, University of Southampton, Southampton, UK;
Yasinska V	Department of Medicine Huddinge, Karolinska Institutet, and Department of Respiratory Medicine, Karolinska University Hospital, Stockholm, Sweden
Zounemat Kermani, N	Data Science Institute, Imperial College, London, UK

Contributors

Ahmed H, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France;
Aliprantis Antonios, Merck Research Laboratories, Boston, USA;

U-BIOPRED

Allen David, North West Severe Asthma Network, Pennine Acute Hospital NHS Trust, UK
Alving Kjell, Dept Women's & Children's Health, Uppsala University, Uppsala, Sweden
Badorrek P, Fraunhofer ITEM; Hannover, Germany
Balgoma David, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Ballereau S, European institute for Systems Biology and Medicine, University of Lyon, France
Barber Clair, NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, UK;
Batuwitage Manohara Kanangana, Data Science Institute, Imperial College, London, UK
Bautmans An, MSD, Brussels, Belgium
Bedding A, Roche Diagnostics GmbH, Mannheim, Germany
Behndig AF, Umeå University, Umea, Sweden
Beleta Jorge, Almirall S.A., Barcelona, Spain;
Berglind A, MSD, Brussels, Belgium
Bochenek Grazyna, II Department of Internal Medicine, Jagiellonian University Medical College, Krakow, Poland;
Braun Armin, Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany;
Campagna D, Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy;
Carayannopoulos Leon, Previously at: MSD, USA;
Casaulta C, University Children's Hospital of Bern, Switzerland
Chaiboonchoe A, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France;
Chaleckis Romanas, Centre of Allergy Research, Karolinska Institutet, Stockholm, Sweden
Davison Timothy Janssen R&D, LLC, Spring House, PA, USA
De Alba Jorge, Almirall S.A., Barcelona, Spain;
De Lepeleire Inge, MSD, Brussels, BE
Dekker Tamara, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Delin Ingrid, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Dennison P, NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences, NIHR-Wellcome Trust Clinical Research Facility, Faculty of Medicine, University of Southampton, Southampton, UK;
Dijkhuis Annemiek, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Dodson Paul, AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden
Draper Aleksandra, BioSci Consulting, Maasmechelen, Belgium;
Dyson K, CROMSOURCE; Stirling, UK
Edwards Jessica, Asthma UK, London, UK;
El Hadjam L, European Institute for Systems Biology and Medicine, University of Lyon
Emma Rosalia, Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy;
Ericsson Magnus, Karolinska University Hospital, Stockholm, Sweden

U-BIOPRED

Faulenbach C, Fraunhofer ITEM; Hannover, Germany
Flood Breda, European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium
Galffy G, Semmelweis University, Budapest, Hungary;
Gallart Hector, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Garissi D, Global Head Clinical Research Division, CROMSOURCE, Italy
Gent J, Royal Brompton and Harefield NHS Foundation Trust, London, UK;
Gerhardsson de Verdier M, AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden
Gibeon D, National Heart and Lung Institute, Imperial College, London, UK;
Gomez Cristina, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Gove Kerry, NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, UK;
Gozzard Neil, UCB, Slough, UK;
Guillmant-Farry E, Royal Brompton Hospital, London, UK
Henriksson E, Karolinska University Hospital & Karolinska Institutet, Stockholm, Sweden
Hewitt Lorraine, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Hoda U, Imperial College, London, UK
Hu Richard, Amgen Inc. Thousand Oaks, USA
Hu Sile, National Heart and Lung Institute, Imperial College, London, UK;
Hu X, Amgen Inc.; Thousand Oaks, USA
Jeyasingham E, UK Clinical Operations, GSK, Stockley Park, UK
Johnson K, Centre for respiratory medicine and allergy, Institute of Inflammation and repair, University Hospital of South Manchester, NHS Foundation Trust, Manchester, UK
Jullian N, European Institute for Systems Biology and Medicine, University of Lyon
Kamphuis Juliette, Longfonds, Amersfoort, The Netherlands;
Kennington Erika J., Asthma UK, London, UK;
Kerry Dyson, CromSource, Stirling, UK;
Kerry G, Centre for respiratory medicine and allergy, Institute of Inflammation and repair, University Hospital of South Manchester, NHS Foundation Trust, Manchester, UK
Klüglich M, Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Knobel Hugo, Philips Research Laboratories, Eindhoven, The Netherlands;
Knox Alan J, Respiratory Research Unit, University of Nottingham, Nottingham, UK;
Kolmert Johan, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Konradsen J R, Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Kots Maxim, Chiesi Pharmaceuticals, SPA, Parma, Italy;
Kretsos Kosmas, UCB, Slough, UK
Krueger L, University Children's Hospital Bern, Switzerland
Kuo Scott, National Heart and Lung Institute, Imperial College, London, UK;
Kupczyk Maciej, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Lambrecht Bart, University of Gent, Gent, Belgium;

U-BIOPRED

Lantz A-S, Karolinska University Hospital & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Larminie Christopher, GSK, London, UK
Larsson L X, AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden
Latzin P, University Children's Hospital of Bern, Bern, Switzerland
Lazarinis N, Karolinska University Hospital & Karolinska Institutet, Stockholm, Sweden
Lefaudeux Diane, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France;
Lemonnier N, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France
Li Chuan-Xing, Respiratory Medicine Unit, Department of Medicine Solna and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; and Department of Respiratory Medicine and Allergy, Karolinska University Hospital Solna, Stockholm, Sweden;
Lone-Latif Saeeda, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Lowe L A, Centre for respiratory medicine and allergy, Institute of Inflammation and repair, University Hospital of South Manchester, NHS Foundation Trust, Manchester, UK
Manta Alexander, Roche Diagnostics GmbH, Mannheim, Germany
Marouzet Lisa, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Martin Jane, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Mathon Caroline, Centre of Allergy Research, Karolinska Institutet, Stockholm, Sweden
McEvoy L, University Hospital, Department of Pulmonary Medicine, Bern, Switzerland
Meah Sally, National Heart and Lung Institute, Imperial College, London, UK;
Menzies-Gow A, Royal Brompton and Harefield NHS Foundation Trust, London, UK;
Metcalf Leanne, Previously at: Asthma UK, London, UK;
Meiser Andrea, Data Science Institute, Imperial College, London, UK
Mikus Maria, Science for Life Laboratory & The Royal Institute of Technology, Stockholm, Sweden;
Monk Philip, Synairgen Research Ltd, Southampton, UK;
Mores N, Università Cattolica del Sacro Cuore, Milan, Italy;
Naz Shama, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Nething K, Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Nicholas Ben, University of Southampton, Southampton, UK
Nihlén U, *Previously* AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden
Nilsson Peter, Science for Life Laboratory & The Royal Institute of Technology, Stockholm, Sweden;
Niven R, North West Severe Asthma Network, University Hospital South Manchester, UK
Nordlund B, Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Nsubuga S, Royal Brompton Hospital, London, UK
Pacino Antonio, Lega Italiano Anti Fumo, Catania, Italy;
Palkonen Susanna, European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium.

U-BIOPRED

Pahus L, Assistance publique des Hôpitaux de Marseille, Clinique des bronches, allergies et sommeil Espace Éthique Méditerranéen, Aix-Marseille Université, Marseille, France
Pellet J, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France
Pennazza Giorgio, Unit of Electronics for Sensor Systems, Department of Engineering, Campus Bio-Medico University of Rome, Rome, Italy
Petrén Anne, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Pink Sandy, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Pison C, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France
Rahman-Amin Malayka, Previously at: Asthma UK, London, UK*;*
Ravanetti Lara, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Ray Emma, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Reinke Stacey, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Reynolds Leanne, Previously at: Asthma UK, London, UK;
Riemann K, Boehringer Ingelheim Pharma GmbH & Co. KG; Biberach, Germany
Robberechts Martine, MSD, Brussels, Belgium
Rocha J P, Royal Brompton and Harefield NHS Foundation Trust
Rossios C, National Heart and Lung Institute, Imperial College, London, UK;
Russell Kirsty, National Heart and Lung Institute, Imperial College, London, UK;
Rutgers Michael, Longfonds, Amersfoort, The Netherlands;
Santini G, Università Cattolica del Sacro Cuore, Milan, Italy;
Santonico Marco, Unit of Electronics for Sensor Systems, Department of Engineering, Campus Bio-Medico University of Rome, Rome, Italy
Saqi M, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France
Schoelch Corinna, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
Scott S, North West Severe Asthma Network, Countess of Chester Hospital, UK
Sehgal N, North West Severe Asthma Network; Pennine Acute Hospital NHS Trust
Selby A, NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences and Human Development and Health, Southampton, UK;
Sjödin Marcus, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Smids Barbara, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Smith Caroline, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK
Smith Jessica, Asthma UK, London, UK;
Smith Katherine M., University of Nottingham, UK;
Söderman P, Dept. Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
Sogbesan A, Royal Brompton and Harefield NHS Foundation Trust, London, UK;
Spycher F, University Hospital Department of Pulmonary Medicine, Bern, Switzerland
Staykova Doroteya, University of Southampton, Southampton, UK

U-BIOPRED

Stephan S, Centre for respiratory medicine and allergy, Institute of Inflammation and repair, University Hospital of South Manchester, NHS Foundation Trust, Manchester, UK
Stokholm J, University of Copenhagen and Danish Pediatric Asthma Centre Denmark
Strandberg K, Karolinska University Hospital & Karolinska Institutet, Stockholm, Sweden
Sunther M, Centre for respiratory medicine and allergy, Institute of Inflammation and repair, University Hospital of South Manchester, NHS Foundation Trust, Manchester, UK
Szentkereszty M, Semmelweis University, Budapest, Hungary;
Tamasi L, Semmelweis University, Budapest, Hungary;
Tariq K, NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences, NIHR-Wellcome Trust Clinical Research Facility, Faculty of Medicine, University of Southampton, Southampton, UK;
Thörngren John-Olof, Karolinska University Hospital, Stockholm, Sweden
Thorsen Jonathan, COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark;
Valente S, Università Cattolica del Sacro Cuore, Milan, Italy;
van Aalderen W M, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands
van de Pol Marianne, Academic Medical Centre, University of Amsterdam, Amsterdam ,The Netherlands;
van Drunen C M, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;
Van Eyll Jonathan, UCB, Slough, UK
Versnel Jenny, Previously at: Asthma UK, London, UK;
Vink Anton, Philips Research Laboratories, Eindhoven, The Netherlands;
von Garnier C, University Hospital Bern, Switzerland;
Vyas A, North west Severe Asthma Network, Lancashire Teaching Hospitals NHS Trust, UK
Wagener A H, Academic Medical Center Amsterdam, Amsterdam, The Netherlands
Wald Frans, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
Walker Samantha, Asthma UK, London, UK;
Ward Jonathan, Histochemistry Research Unit, Faculty of Medicine, University of Southampton, Southampton, UK;
Wetzel Kristiane, Boehringer Ingelheim Pharma GmbH, Biberach, Germany
Wiegman Coen, National Heart and Lung Institute, Imperial College, London, UK;
Weiszhart Z, Semmelweis University, Budapest, Hungary
Williams Siân, International Primary Care Respiratory Group, Aberdeen, Scotland;
Yang Xian, Data Science Institute, Imperial College, London, UK
Yeyasingham Elizabeth, UK Clinical Operations, GSK, Stockley Park, UK;
Yu W, Amgen Inc.; Thousand Oaks, USA
Zetterquist W, Dept. Women's and Children's Health & Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden
Zolkipli Z, NIHR Southampton Respiratory Biomedical Research Unit, Clinical and Experimental Sciences and Human Development and Health, Southampton, UK;

U-BIOPRED

Zwinderman A H, Academic Medical Centre, University of Amsterdam, The Netherlands;

Partner organisations	
Novartis Pharma AG	University of Southampton, Southampton, UK
Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands	Imperial College London, London, UK
University of Catania, Catania, Italy	University of Rome 'Tor Vergata', Rome, Italy
Hvidore Hospital, Hvidore, Denmark	Jagiellonian Univ. Medi.College, Krakow, Poland
University Hospital, Inselspital, Bern, Switzerland	Semmelweis University, Budapest, Hungary
University of Manchester, Manchester, UK	Université d'Aix-Marseille, Marseille, France
Fraunhofer Institute, Hannover, Germany	University Hospital, Umea, Sweden
Ghent University, Ghent, Belgium	Ctr. Nat. Recherche Scientifique, Lyon, France
Università Cattolica del Sacro Cuore, Rome, Italy	University Hospital, Copenhagen, Denmark
Karolinska Institutet, Stockholm, Sweden	Nottingham University Hospital, Nottingham, UK
University of Bergen, Bergen, Norway	Netherlands Asthma Foundation, Leusden, NL
European Lung Foundation, Sheffield, UK	Asthma UK, London, UK
European. Fed. of Allergy and Airways Diseases Patients' Associations, Brussels, Belgium	Lega Italiano Anti Fumo, Catania, Italy
International Primary Care Respiratory Group, Aberdeen, Scotland	Philips Research Laboratories, Eindhoven, NL
Synairgen Research Ltd, Southampton, UK	Aerocrine AB, Stockholm, Sweden
BioSci Consulting, Maasmechelen, Belgium	Almirall
AstraZeneca BioPharmaceuticals R&D	Boehringer Ingelheim
Chiesi	GlaxoSmithKline
Roche	UCB
Janssen Biologics BV	Amgen NV
Merck Sharp & Dome Corp	

MEMBERS OF THE ETHICS BOARD			
Name	**Task**	**Affiliation**	**e-mail**
Jan-Bas Prins	Biomedical research	LUMC/the Netherlands	J.B.Prins@lumc.nl
Martina Gahlemann	Clinical care	BI/Germany	Martina.Gahlemann@boehringer-ingelheim.com
Luigi Visintin	Legal affairs	LIAF/Italy	visintin@inrete.it
Hazel Evans	Paediatric care	Southampton/UK	hazel.evans@uhs.nhs.uk
Martine Puhl	Patient representation (co-chair)	NAF/ the Netherlands	martine@puhl.nl
Lina Buzermaniene	Patient representation	EFA/Lithuania	lina.buzermaniene@pavb.lt
Val Hudson	Patient representation	Asthma UK	hudsonval7@gmail.com
Laura Bond	Patient representation	Asthma UK	lvbond22@googlemail.com
Pim de Boer	Patient representation and pathobiology	IND	deboer.pim@hetnet.nl

Guy Widdershoven	Research ethics	VUMC/the Netherlands	g.widdershoven@vumc.nl
Ralf Sigmund	Research methodology and biostatistics	BI/Germany	ralf.sigmund@boehringer-ingelheim.com

THE PATIENT INPUT PLATFORM	
Name	**Country**
Amanda Roberts	UK
David Supple (chair)	UK
Dominique Hamerlijnck	The Netherlands
Jenny Negus	UK
Juliëtte Kamphuis	The Netherlands
Lehanne Sergison	UK
Luigi Visintin	Italy
Pim de Boer (co-chair)	The Netherlands
Susanne Onstein	The Netherlands

MEMBERS OF THE SAFETY MONITORING BOARD	
Name	**Task**
William MacNee	Clinical care
Renato Bernardini	Clinical pharmacology
Louis Bont	Paediatric care and infectious diseases
Per-Ake Wecksell	Patient representation
Pim de Boer	Patient representation and pathobiology (chair)
Martina Gahlemann	Patient safety advice and clinical care (co-chair)
Ralf Sigmund	Bio-informatician